

Análisis de los Sets de Datos

Luciano Hernández

2025-11-25

Contents

Extraer 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.

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 (67.82%) versus 2,017 en zonas rurales (32.18%)

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

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

Optimizaciones Realizadas

- **Unificación temporal:** Fusión de las columnas “Fecha” y “Hora” en un único campo datetime
- **Normalización de datos:** Conversión de 8 variables de texto a tipo factor para mejorar el procesamiento estadístico:
 - “Región” -“Comuna”
 - “Tipo Accidente”
 - “Tipo (CONASET)”
 - “Zona”
 - “Ubicación Relativa”
 - “Causa (CONASET)”
 - “Causa Accidente”

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

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

Calidad de Señal GPS

- **Valor fix:** Constante en 3 para todos los registros, sugiriendo un posicionamiento 3D consistente
- **Precisión:** Oscila entre 121 y 158 unidades (media: 125.9, mediana: 125)

Aspectos Requieren Clarificación

Variables por Definir

- **cts:** Propósito y unidades no especificadas
- **fix:** Escala de calidad de posicionamiento GPS no documentada
- **Unidades de precisión:** Métrica de exactitud no especificada
- **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

Normalización Temporal

- Fusión de columnas `date` y `hora` en un campo datetime unificado

Corrección de Coordenadas Geográficas

Función aplicada:

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

Objetivo: Ajuste de coordenadas chilenas mediante reposicionamiento del punto decimal basado en el orden de magnitud

Normalización de Velocidad 2D

Función aplicada:

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

Objetivo: Alineación de órdenes de magnitud entre velocidades 2D y 3D mediante escalado decimal dinámico

Observaciones Técnicas

Los ajustes aplicados sugieren que el sistema de captura original podría presentar inconsistencias en el formato de salida numérico, requiriendo post-procesamiento para una interpretación correcta de las mediciones GPS.

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

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

Distribución Direccional

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

Métricas de Desempeño Técnico

Calidad de Señal GPS

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

Velocidades Operacionales

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

Estados del Sistema

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

Procesamiento y Normalización de Datos

Estructuración Multi-carpetas

- **Estrategia:** Procesamiento paralelo de 31 unidades operativas
- **Identificación:** Incorporación de metadatos de carpeta origen
- **Consolidación:** Unificación en dataset coherente

Transformaciones Aplicadas

Normalización Temporal

- **Timestamp Chile:** Conversión a formato POSIXct
- **Timestamp UTC:** Sincronización con estándar internacional
- **Zona horaria:** Conservación de huso local (America/Santiago)

Clasificación de Calidad de Señal Escala DOP (Dilution of Precision):

- **Excelente:** $DOP < 1$
- **Muy Buena:** $1 \leq DOP < 2$
- **Buena:** $2 \leq DOP < 5$
- **Regular:** $5 \leq DOP \leq 10$
- **Mala:** $DOP > 10$

Optimización de Tipos de Datos

- **Coordenadas:** Conversión a numérico con estandarización decimal
- **Identificadores:** Transformación a factores para análisis categórico
- **Distancias:** Normalización de formatos numéricos

Variables Críticas Procesadas

- `EVT_GPS_TIME_CHILE_STR, EVT_GPS_TIME_UTC_0` -> POSIXct
- `EVT_SERVICIO_ID, EVT_PPU` -> Factor
- `Y_EVT_GPS_LAT, X_EVT_GPS_LON` -> Numérico
- `EVT_DISTANCIA_RECORRIDA` -> Numérico
- `EVT_SEÑAL` -> Factor (derivado de DOP)

Análisis por Unidad Operativa

Patrones de Operación por Unidad

Frecuencia de Reporte

- **Típico:** 1:00-1:03 minutos entre registros por vehículo
- **Variaciones:** UN52_265 muestra intervalos más irregulares (3.5s - 1min)
- **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)
- **UN22_257:** Mayor variabilidad en intervalos (59s - 22:31 min)
- **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

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)

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

Optimizaciones Aplicadas

- **Unificación temporal:** Fusión de campos “FECHA” y “HORA” en timestamp único
- **Normalización de tipos de datos:**
 - **Conversión a factor:** ZONA, COMUNA, TIPO, CAUSA, SECTOR, TRIBUNAL
 - **Conversión a numérico:** KM (donde aplicable) y PARTE

Creación de Categorías Agrupadas

- **Variable “MACRO TIPO”:**
 - **CHOQUE:** Agrupa “CHOQUE”, “CHOQUE FRONTAL”, “CHOQUE LATERAL”, “CHOQUE POSTERIOR”
 - **COLISION:** Agrupa “COLISION”, “COLISION FRONTAL”, “COLISION LATERAL”, “COLISION PERPENDICULAR”, “COLISION POR ALCANCE”
 - **Conserva valores originales** para categorías no agrupables

Observaciones Relevantes

La concentración temporal del dataset (solo 30 días) permite un análisis detallado de patrones de siniestralidad en un período específico, aunque limita la identificación de tendencias estacionales. La predominancia de accidentes en zonas urbanas refleja la distribución poblacional y de tráfico vehicular del país.

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 (año no especificado)
- **Cobertura temporal:** 37 días de monitoreo continuo

Clasificación de Incidentes

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

Tipología de Incidentes

Distribución por categoría:

- Congestión: 191 casos
- Obstrucciones en la vía: 38 casos
- Condición de tráfico: 26 casos
- Problema Infraestructura de control de tránsito: 25 casos
- Accidentes de tránsito: 20
- Intervenciones de emergencia: 7 casos
- Otros: 5 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

Aspectos Requieren Clarificación

Información Crítica Faltante

- **Año de referencia:** No especificado en los datos originales
- **Categorías direccionales:** Necesidad de confirmar si “N-S, S-N” y “S-N, N-S” representan categorías distintas o requieren unificación

Procesamiento y Normalización de Datos

Extracción y Estructuración

- **Identificación de hojas relevantes:** Filtrado por nombres con formato fecha (dd-mm)
- **Consolidación temporal:** Unificación de 37 hojas diarias en dataset único
- **Construcción de timeline:** Asignación de fechas basada en nombres de hojas

Limpieza y Estandarización

Normalización de Variables Categóricas:

- “**Reporte completo? (SI/NO)**”:
 - *Valores estandarizados:* “SI” para [“is”, “si”, “SI”, “su”], “NO” para [“no”, “NO”]
 - *Valores excluidos:* “-” convertido a NA
- “**es incidente de tráfico? (si/no)**”:
 - *Valores estandarizados:* “SI” para [“si”, “SI”], “NO” para [“no”, “NO”]
- **Direccionalidad (“dir”):**
 - *Normalización de notaciones:* Unificación de variantes (“e-o” → “E-O”, “n-s” → “N-S”, etc.)
 - *Patrones bidireccionales:* Estandarización de formatos compuestos
 - *Valores inválidos:* [“-”, “n/a”, “n/A”] convertidos a NA

Sistema de Clasificación por Color

Mapeo cromático de tipos de incidente:

- FF00B0F0 → Intervenciones de Emergencia
- FF00FFFF, FFFFFFFF00, NA → Congestión
- FFFF0000 → Accidente de tránsito
- FFFF0OFF → Obstrucciones en la vía
- FFFF6600 → Problema Infraestructura de control
- FFFFC000 → Condición de tráfico
- FF00B050 → Otros

Optimizaciones Geográficas y Temporales

- **Coordenadas:** Separación de campo “coord” en latitud/longitud numéricas
- **Timestamp:** Fusión de fecha y hora en campo datetime unificado
- **Completitud:** Rellenado de horas faltantes con valores anteriores
- **Estructura de datos:** Conversión de variables clave a tipo factor

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:** 09:06 a 17:58 horas (fecha no especificada)
- **Duración:** Aproximadamente 9 horas de monitoreo continuo

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

Aspectos que Requieren Clarificación

Información Crítica Faltante

- **Fecha de medición:** No especificada en los datos originales
- **Clasificación vehicular:** Necesidad de confirmar si vehículos sin comentario deben categorizarse automáticamente como “Auto”

Procesamiento y Normalización de Datos

Estructuración Multi-hoja

- **Organización por dirección:** Cada hoja representa una dirección de flujo (Concepción→Hualqui / Hualqui→Concepción)
- **Consolidación:** Unificación de todas las hojas en dataset único
- **Metadatos direccionales:** Incorporación de columna “dirección” desde nombres de hojas

Limpieza y Estandarización

Normalización de Localización

- **Unificación geográfica:** Campo “Lugar (dirección)” consolidado por hoja
- **Eliminación de redundancias:** Remoción de 9 columnas vacías o duplicadas (...5 a ...13)

Procesamiento Temporal

- **Compleitud horaria:** Rellenado de valores faltantes usando observaciones anteriores
- **Formato estandarizado:** Conversión a tipo hms para análisis temporal

Estandarización de Categorías

Comentarios vehiculares:

- **Maniobras:** Unificación de variantes [“auto, doble a Lo Plaza”, “Auto, doble a Lo Plaza”, “Auto, gira a Lo Plaza”] → “Auto, dobló a Lo Plaza”
- **Tipos vehiculares:** Normalización de [“camion”, “Camion”] → “Camión” y [“micro”, “Micro”] → “Micro”

Clasificación Vehicular Sistematizada

Variable “Vehículo”:

- **Valor por defecto:** “Auto” para registros sin comentario o con referencias a automóviles
- **Especificación:** Conservación de categorías explícitas [“Camión”, “Micro”, “MOTO”, “retroexcavadora”]
- **Optimización:** Conversión a tipo factor para análisis categórico

Optimizaciones Finales

- **Estructuración:** Conversión de variables clave a tipo factor
- **Consistencia:** Validación de integridad de datos numéricos
- **Preparación analítica:** Dataset listo para análisis estadístico y temporal

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:** 08:08 a 17:58 horas (fecha no especificada)
- **Duración:** Aproximadamente 10 horas de monitoreo continuo

Análisis de Velocidades

- **Rango de velocidades:** 5 a 23 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

Aspectos que Requieren Clarificación

Información Crítica Faltante:

- **Fecha de medición:** No especificada en los datos originales
- **Contexto temporal:** Día de semana vs. fin de semana no identificado

Procesamiento y Normalización de Datos

Estrategia de Extracción Multi-hoja:

- **Filtrado inteligente:** Identificación de hojas relevantes mediante patrones “21” y “mayo”
- **Organización direccional:** Cada hoja representa un sentido de flujo específico
- **Consolidación estructurada:** Unificación de múltiples hojas en dataset coherente

Corrección de Formatos Temporales

Normalización de Horarios

- **Problema identificado:** Formato decimal incorrecto en celdas horarias (ej: “8,15” en lugar de “8:15”)
- **Solución aplicada:**
 - Conversión de formato decimal a fracción diaria
 - Reconstrucción de horas y minutos mediante operaciones matemáticas
 - Formateo consistente a “HH:MM:SS”

Completitud de Series Temporales

- **Rellenado inteligente:** Propagación de valores horarios hacia adelante
- **Consistencia temporal:** Garantía de continuidad en las mediciones

Estandarización de Categorías y Comentarios

Unificación de Términos de Congestión

Agrupamiento por patrones:

- *Formación de colas:* [“cola por funeral”, “forman cola”] → “Forman cola”
- *Incorporación a colas:* Variantes de “llega a cola” unificadas
- *Agrupamiento vehicular:* Múltiples expresiones de “llegan juntos” estandarizadas

Clasificación Vehicular

- *Transporte público:* [“micro”, “Micro”] → “Micro”
- *Motocicletas:* [“moto”, “Moto”] → “Moto”

Optimizaciones Estructurales

- **Geolocalización unificada:** Consolidación de “Lugar.(dirección)” por hoja
- **Metadatos direccionales:** Incorporación sistemática desde nombres de hojas
- **Tipificación eficiente:** Conversión a factores para análisis categórico
- **Limpieza de datos:** Eliminación de registros sin medición de velocidad

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 (66 minutos)
- **Según timestamp del GPS:** 31 de diciembre de 2024, 23:59:22 - 1 de enero de 2025, 01:06:49 (67 minutos)
- **Discrepancia temporal:** Diferencia de sincronización entre sistemas de aproximadamente 1 minuto

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

Corrección de Formato de Archivo

- **Problema original:** Formato CSV con delimitadores inconsistentes y caracteres de escape
- **Transformación aplicada:**
 - Reemplazo de “;” por espacio para unificación
 - Sustitución de , por ; como nuevo delimitador
 - Eliminación de comillas redundantes
- **Resultado:** Estructura tabular legible para importación

Optimización de Estructura de Datos

- **Eliminación de columnas vacías:** `ubicacion_` y `velocidad_` removidas por contener exclusivamente valores nulos
- **Normalización temporal:**
 - `fecha_auto` y `fecha_gps_` convertidas a formato datetime
 - Preservación de precisiones de segundos
- **Tipificación categórica:**
 - `evento_` y `direccion_` convertidas a factor para análisis estadístico

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

Corrección de Estructura de Columnas:

- **Problema original:** Columnas duplicadas “MARCA”, “MODELO” y “NÚMERO DE SERIE” para cámara y codificador
- **Solución aplicada:**
 - Diferenciación clara mediante sufijos “DE LA CÁMARA” y “DEL CODIFICADOR DE VIDEO”
 - Asignación basada en posición original en el dataset (columnas 1-11: cámara, 12-17: codificador)

Optimización para Análisis:

- **Normalización categórica:** Conversión de 9 variables clave a tipo factor:
 - COMUNA
 - ESTADO
 - TIPO DE ENLACE
 - PROVEEDOR
 - MARCA DE LA CÁMARA
 - MODELO DE LA CÁMARA
 - MARCA DEL CODIFICADOR
 - MODELO DEL CODIFICADOR
 - REGIÓN
- **Manejo de valores faltantes:** Identificación explícita de valores “—” como NA

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:** Aproximadamente 11 meses de datos continuos
- **Registros diarios promedio:** ~850 alertas por día

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

Transformación de Coordenadas Geográficas

- **Problema original:** Campo “Location” en formato WKT (Well-Known Text): Point(lng lat)
- **Solución aplicada:**

- Extracción de longitud (lng) mediante regex: (?=<Point\(\)-?\d+\.\d*)
- Extracción de latitud (lat) mediante regex: -?\d+\.\d*(?=\\)\$
- Conversión a tipos numéricos para análisis espacial

Normalización de Fechas

- **Formato original:** Día-Mes-Año con abreviaturas en español
- **Conversión aplicada:**
 - Mapeo de abreviaturas españolas a inglesas: {"ene":"jan", "feb":"feb", ..., "dic":"dec"}
 - Transformación a formato Date usando `dmy()`

Optimización para Análisis

- **Tipificación categórica:** Conversión a factor de 4 variables clave:
 - Country
 - City
 - Type
 - Subtype
- Eliminación de redundancias: Remoción de columna “Location” original
- Estructura final: Dataset con 281759 registros × 8 columnas optimizadas

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 accidentes reportados por día en promedio

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

Procesamiento y Normalización de Datos

Transformación de Georreferenciación

- **Formato original:** Campo “Location” en estándar WKT (Well-Known Text)
- **Metodología de extracción:**

- Longitud (lng): Patrón regex (?<=Point\\() -?\\d+\\.?.?\\d*
- Latitud (lat): Patrón regex -?\\d+\\.?.?\\d*(?=\\\$)
- **Resultado:** Coordenadas decimales para análisis geoespacial avanzado

Estandarización de Fechas

- **Problema identificado:** Nomenclatura de meses en español
- **Solución implementada:**
 - Mapeo sistemático: {"ene": "jan", "feb": "feb", ..., "dic": "dec"}
 - Conversión robusta mediante función dmy()
- **Output:** Timeline consistente para análisis temporal

Optimización de Estructura de Datos

- **Eliminación de redundancias:** Remoción de columna “Location” original
- **Normalización categórica:**
 - Country y City convertidas a tipo factor
 - Habilitación de análisis agregados por ubicación
- **Dataset final:** 281981 registros × 7 columnas optimizadas

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 irregularidades reportadas por día

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

Estandarización Temporal

- **Formato original:** Fechas con nomenclatura de meses en español
- **Transformación aplicada:**
 - Mapeo sistemático: {"ene": "jan", "feb": "feb", ..., "dic": "dec"}
 - Conversión a formato Date mediante dmy()
- **Resultado:** Timeline consistente para análisis diario y estacional

Optimización de Estructura de Datos

- **Manejo de valores nulos:** Identificación explícita de “null” como NA
- **Normalización categórica:**
 - Country, City, Street, Cause convertidas a tipo factor
 - Habilitación de análisis agregados por múltiples dimensiones
- **Dataset final:** 13502 registros \times 8 columnas optimizadas

Discretizar datos

Descripción Matemática

Sea la función:

$$quantum.data : \mathbb{R}^{4 \times n} \times \mathbb{R}^2 \times (\mathcal{P}(\mathbb{R}) \rightarrow \mathbb{R}) \rightarrow \mathbb{R}^{5 \times m}$$

Donde:

- **Dominio:** Matrices de datos $4 \times n$ + parámetros de discretización + función de compilación
- **Codominio:** Matrices de datos discretizados $5 \times m$ con $m \leq n$

Parámetros de Entrada

- $\text{time.data} \in \mathbb{R}^n$: Vector temporal (segundos)
- $\text{lng.data} \in \mathbb{R}^n$: Vector de longitudes (grados)
- $\text{lat.data} \in \mathbb{R}^n$: Vector de latitudes (grados)
- $\text{field.data} \in \mathbb{R}^n$: Vector de valores del campo de interés
- $\text{time.gap} \in \mathbb{R}^+$: Intervalo temporal de discretización (segundos)
- $\text{space.gap} \in \mathbb{R}^+$: Intervalo espacial de discretización (metros)
- $\text{compile.function} : \mathcal{P}(\mathbb{R}) \rightarrow \mathbb{R}$: Función de agregación (ej: media, máximo, mínimo)

Proceso de Discretización

1. Conversión de Unidades Espaciales

Definimos las constantes de conversión:

$$\text{meters.per.degree.lat} \equiv 111320$$

$$\text{meters.per.degree.lng}(x) \equiv 111320 \times \cos\left(x \times \frac{\pi}{180}\right)$$

2. Discretización Temporal

Para $\text{time.gap} > 0$:

- Sea $t_{\min} = \min(\text{time.data})$, $t_{\max} = \max(\text{time.data})$
- Construimos los intervalos: $B_t = t_{\min}, t_{\min} + \Delta t, t_{\min} + 2\Delta t, \dots, t_{\max}$ donde $\Delta t = \text{time.gap}$
- Asignamos cada t_i al bin temporal $b_t(i) = k$ tal que $t_i \in [B_t[k], B_t[k+1]]$

3. Discretización Espacial

Para $\text{space.gap} > 0$:

- Convertimos rangos a metros:
 - $R_{lat}^m = [\min(\text{lat.data}), \max(\text{lat.data})] \times 111320$
 - $R_{lng}^m = [\min(\text{lng.data} \times f_{lng}), \max(\text{lng.data} \times f_{lng})]$ donde $f_{lng} = \text{meters.per.degree.lng}(\text{lat.media})$
- Construimos bins en metros:
 - $B_{lat}^m = R_{lat}^m[1], R_{lat}^m[1] + \Delta s, \dots, R_{lat}^m[2]$
 - $B_{lng}^m = R_{lng}^m[1], R_{lng}^m[1] + \Delta s, \dots, R_{lng}^m[2]$ donde $\Delta s = \text{space.gap}$
- Convertimos bins a grados:
 - $B_{lat} = B_{lat}^m / 111320$
 - $B_{lng} = B_{lng}^m / f_{lng}$

4. Agregación por Celdas

Para cada celda espacio-temporal (b_t, b_{lat}, b_{lng}) :

- $\text{timecelda} = \text{meant}_i : t_i \in b_t, lat_i \in blat, lng_i \in b_{lng}$
- $\text{lngcelda} = \text{meanlng}_i : t_i \in b_t, lat_i \in blat, lng_i \in b_{lng}$
- $\text{latcelda} = \text{meanlat}_i : t_i \in b_t, lat_i \in blat, lng_i \in b_{lng}$
- $\text{fieldcelda} = \text{compile.function} \text{field}_i : t_i \in b_t, lat_i \in blat, lng_i \in b_{lng}$
- $\text{n.pointscelda} = |i : t_i \in b_t, lat_i \in blat, lng_i \in b_{lng}|$

Salida

La función retorna una matriz $\mathbb{R}^{5 \times m}$ con columnas:

1. time: Tiempo representativo de la celda
2. lng: Longitud representativa de la celda
3. lat: Latitud representativa de la celda
4. field: Valor del campo agregado
5. n.points: Número de puntos originales en la celda

Propiedades Matemáticas

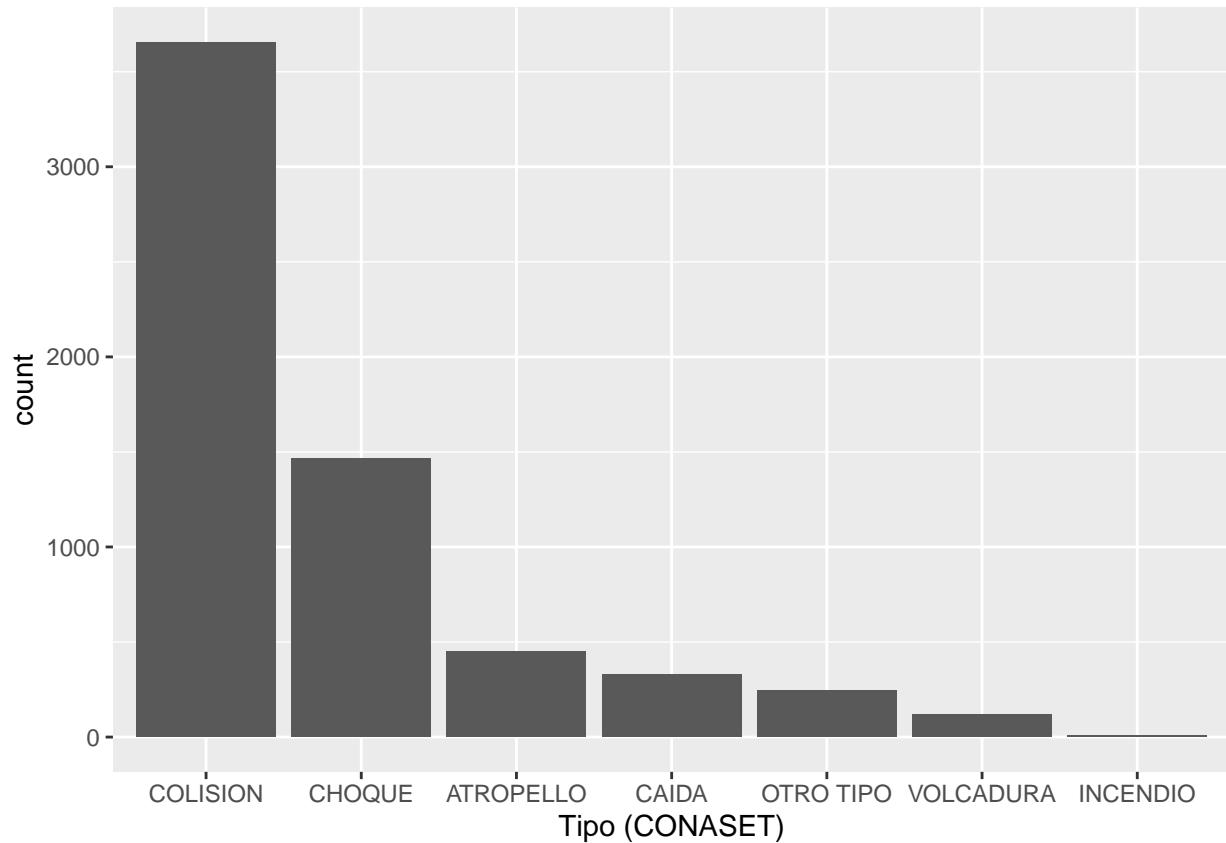
- **Reducción dimensional:** $m \leq n$ (generalmente $m \ll n$)
- **Conservación de información:** Los valores agregados representan estadísticos de los datos originales
- **Adaptabilidad espacial:** La discretización longitudinal considera la variación métrica con la latitud
- **Robustez:** Manejo de valores en los límites mediante asignación a bins adicionales

Esta función implementa esencialmente un **proceso de cuantización vectorial** en el espacio \mathbb{R}^3 (tiempo \times latitud \times longitud) con métricas adaptativas.

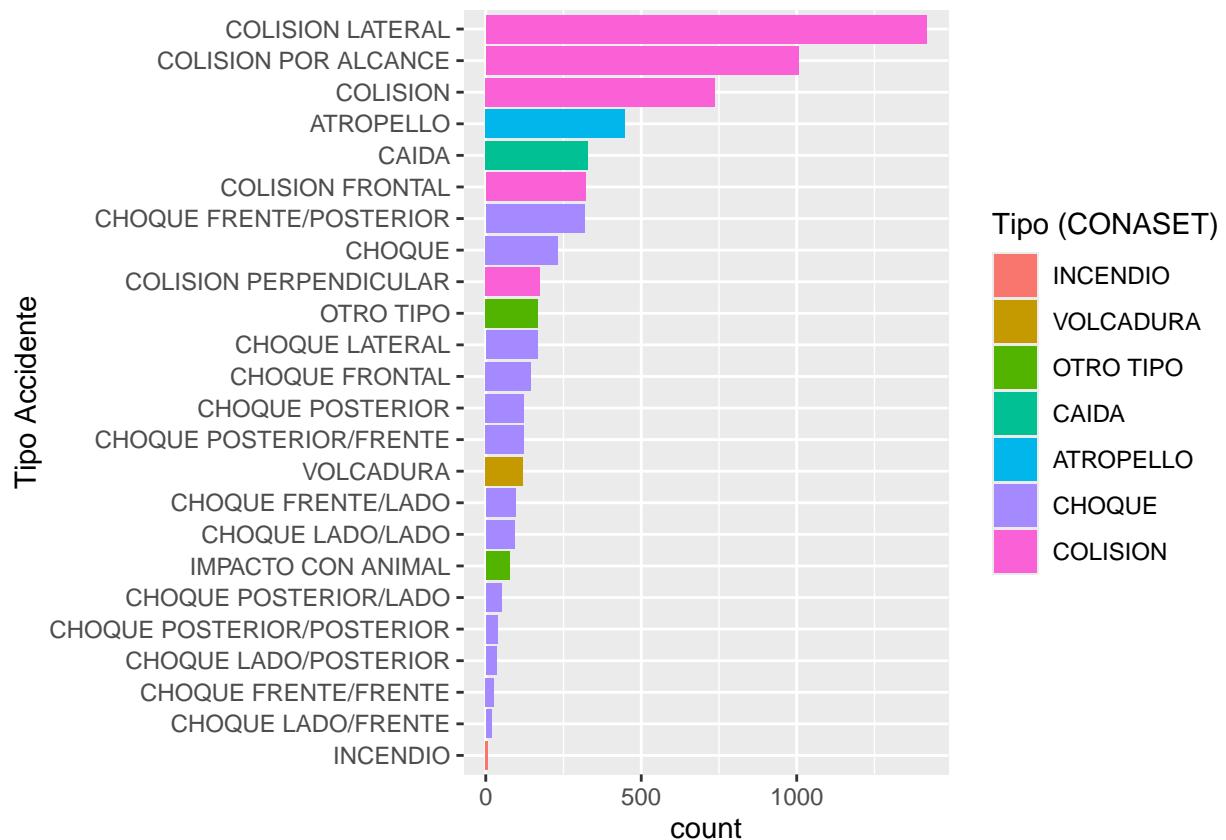
Visualizar datos

Accidentes interurbanos

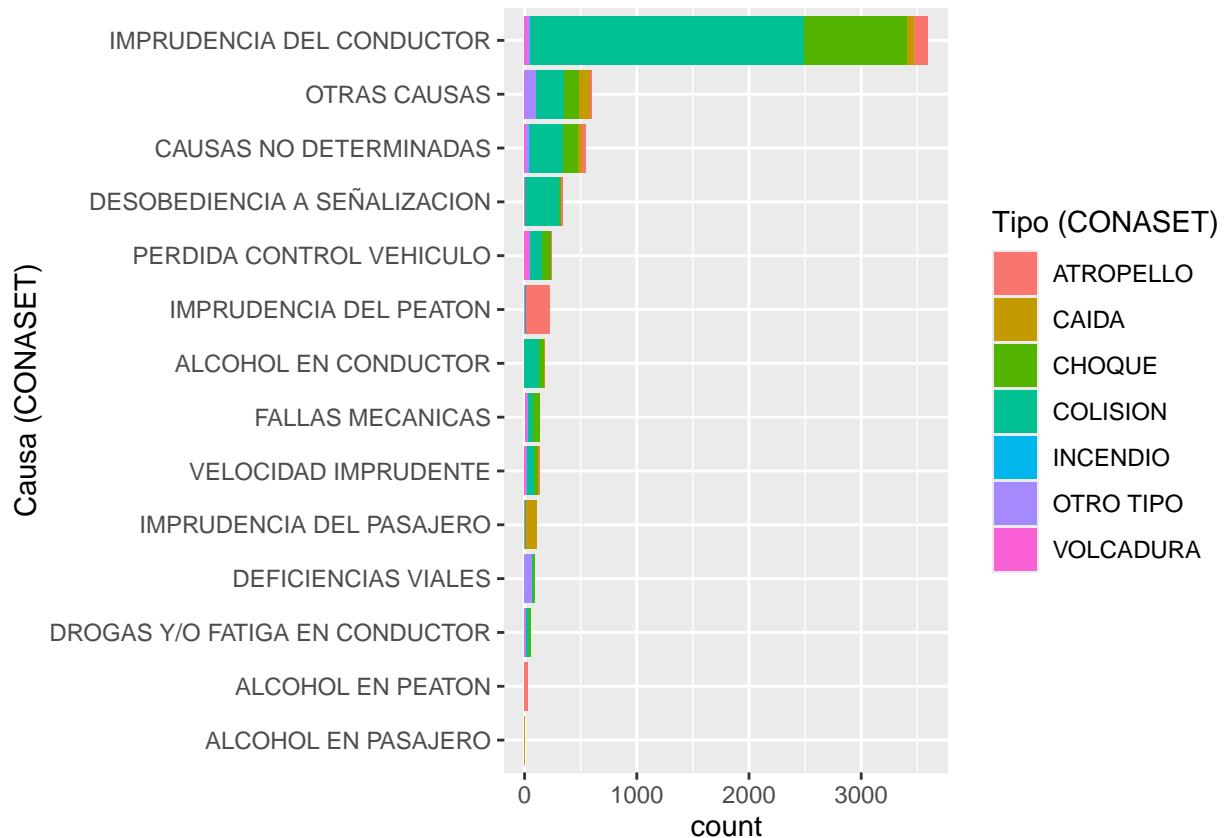
Distribución por tipo de accidente según CONASET



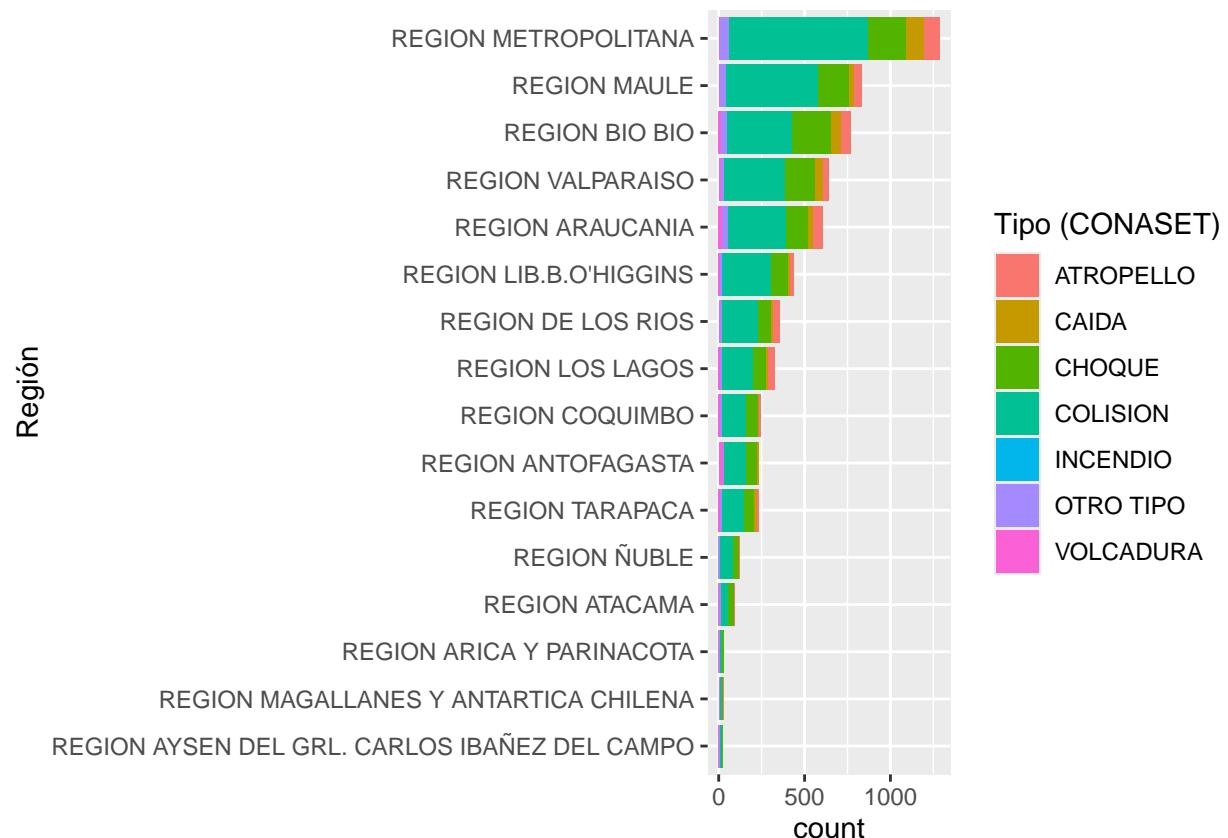
Distribución por tipo de accidente



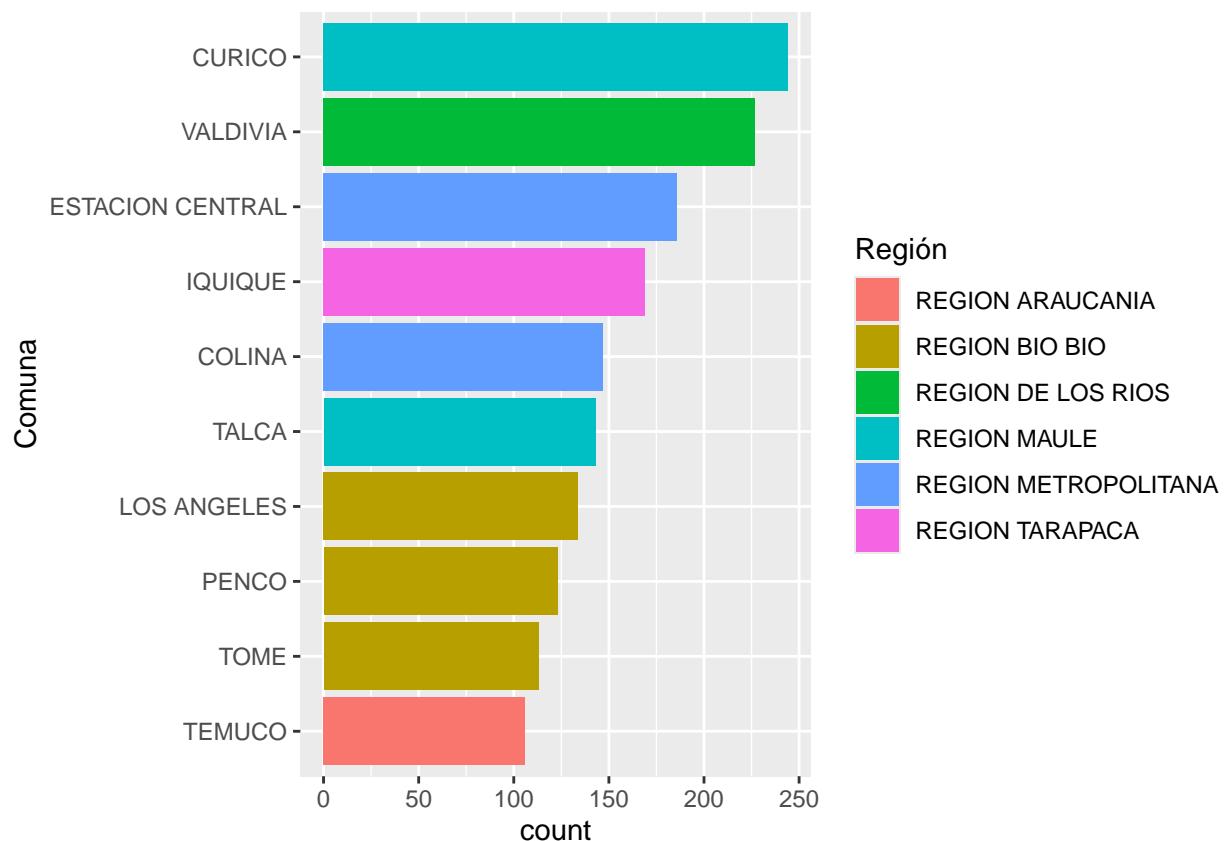
Distribución por causa del accidente según CONASET



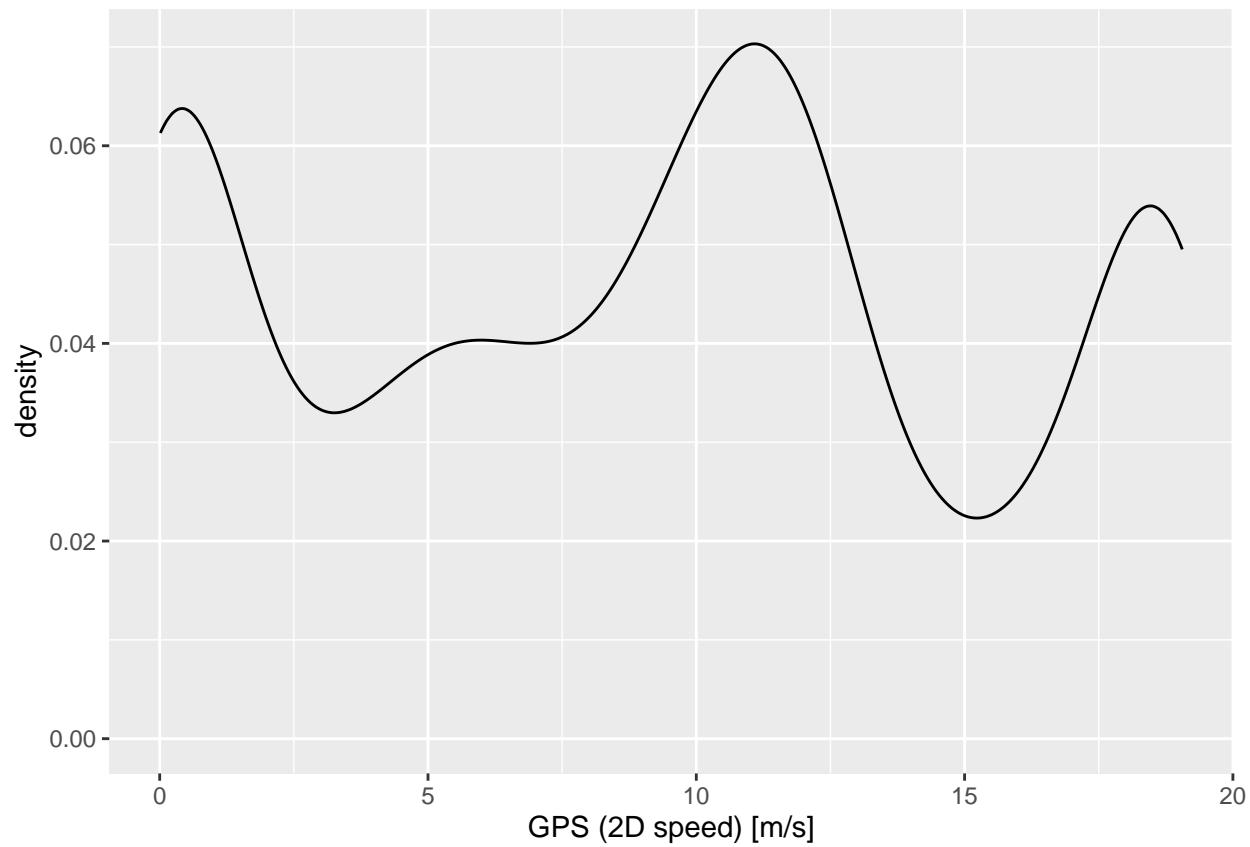
Distribución por región del accidente



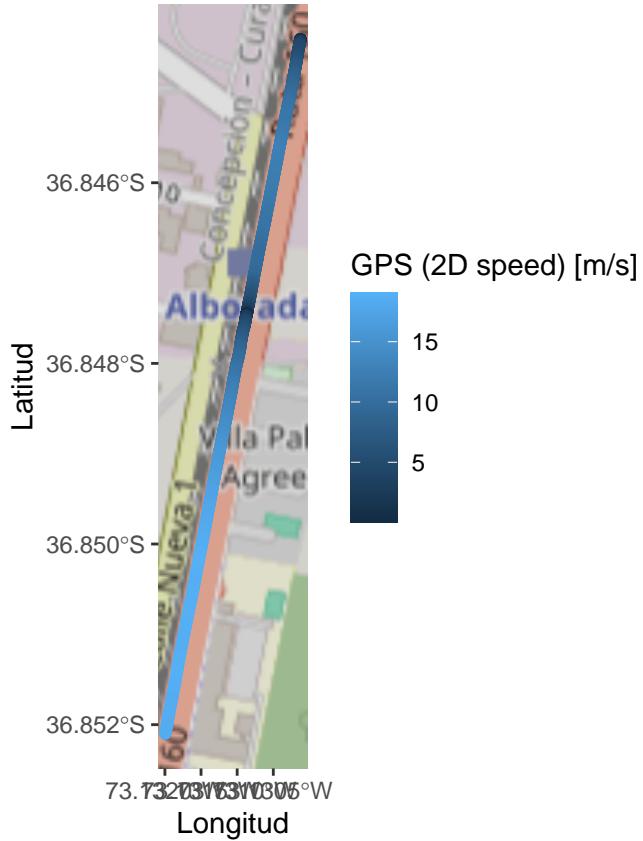
10 comunas con más accidentes



Mediciones de velocidad por satélite



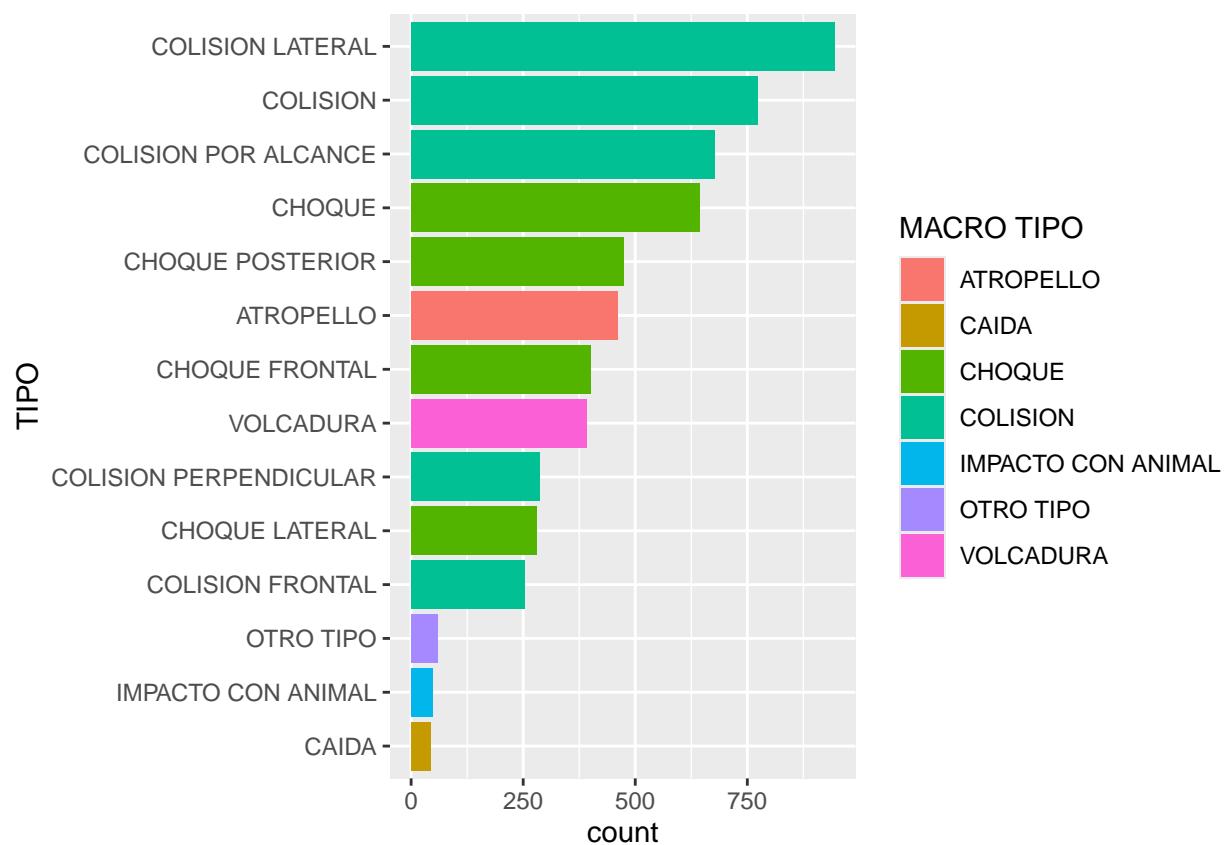
```
## Zoom: 15
```



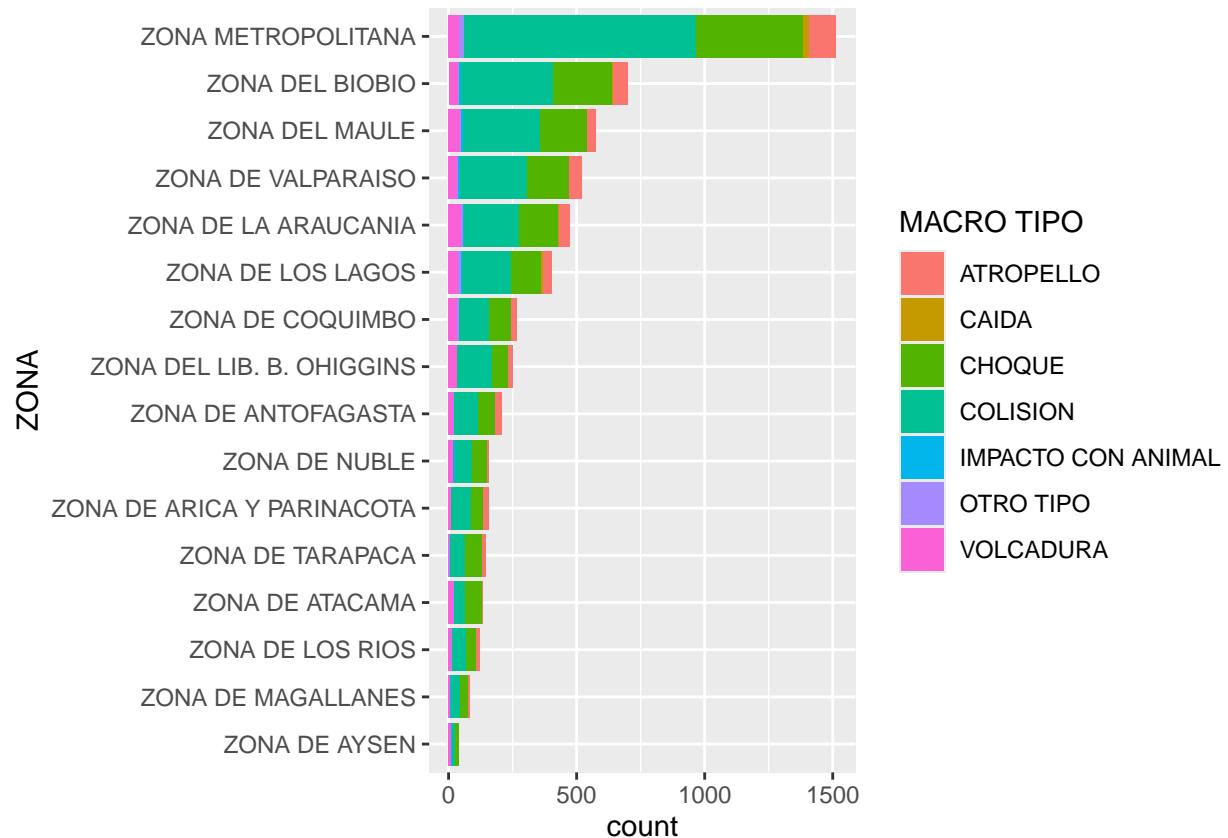
Los datos son de diferentes puntos de la Ruta 160 de San Pedro de la Paz

Accidentes reportados por carabineros

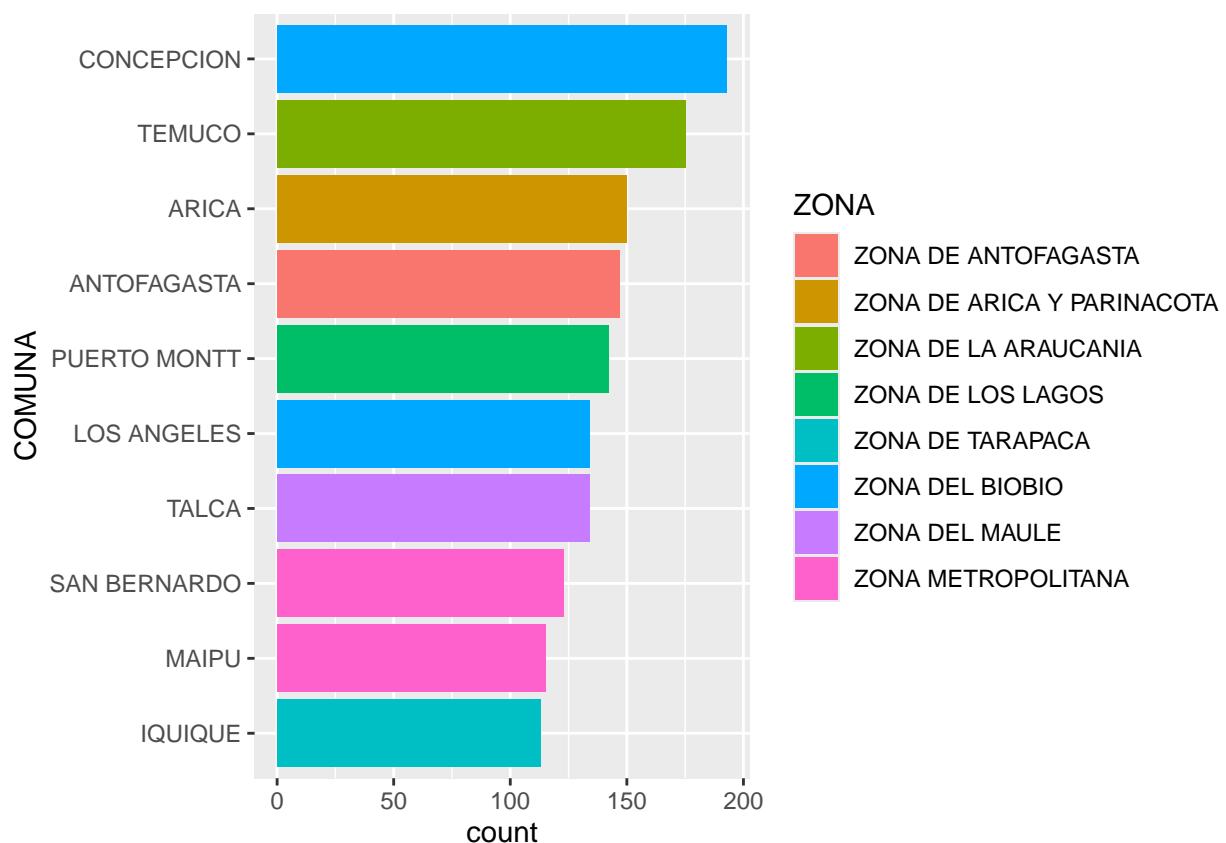
Distribución por tipo



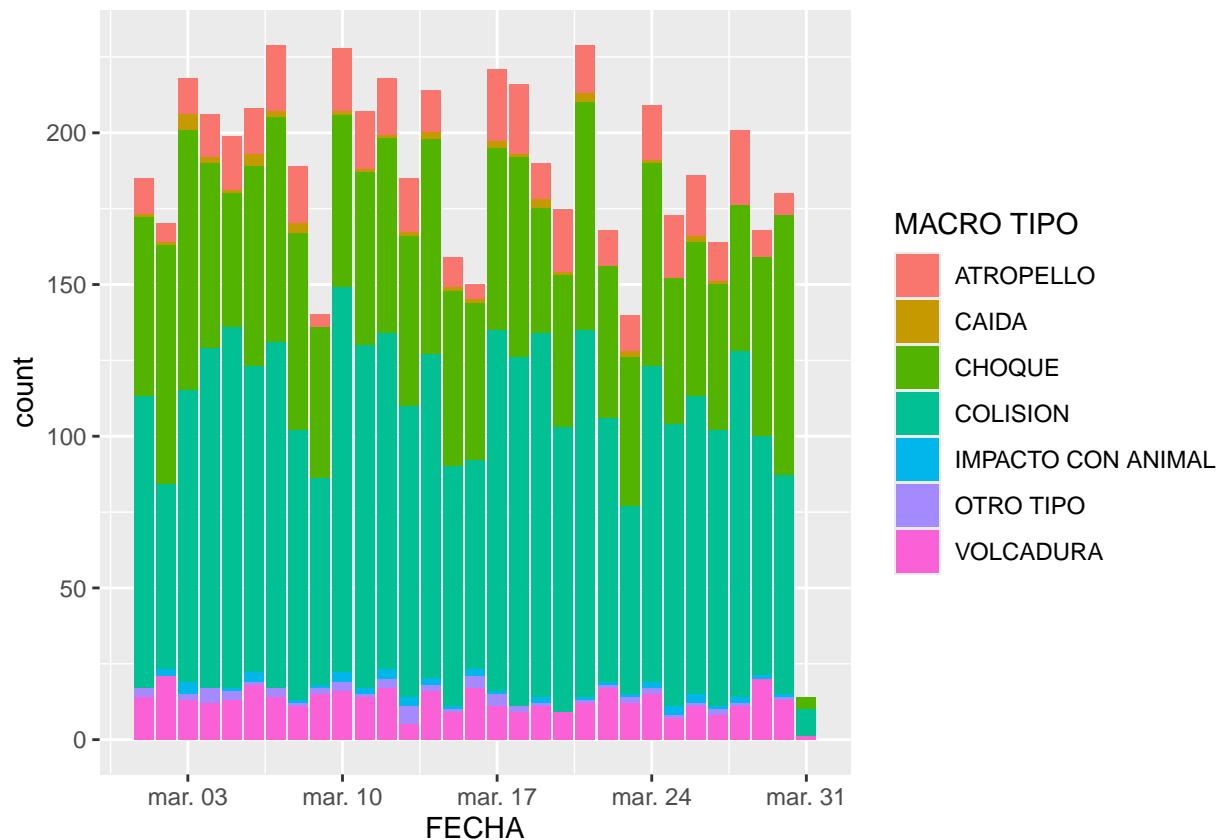
Distribución por región



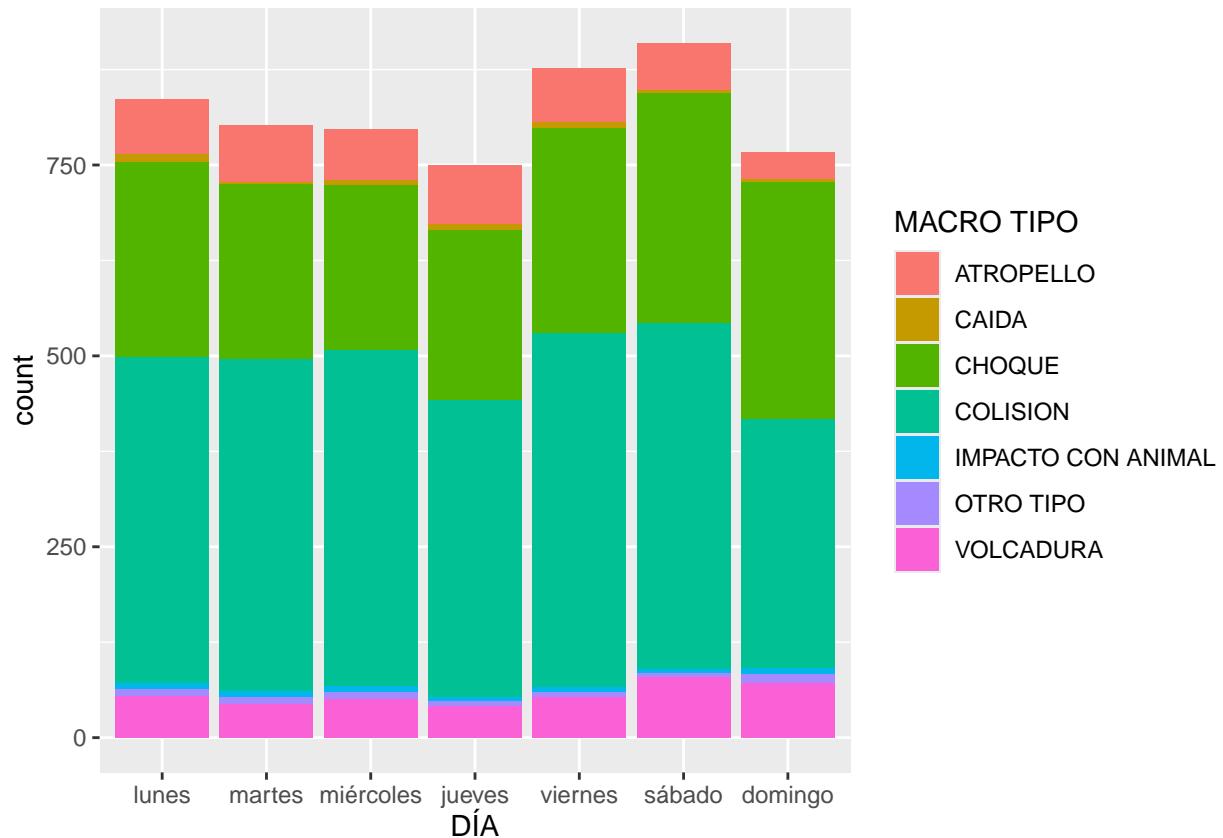
10 comunas con más accidentes



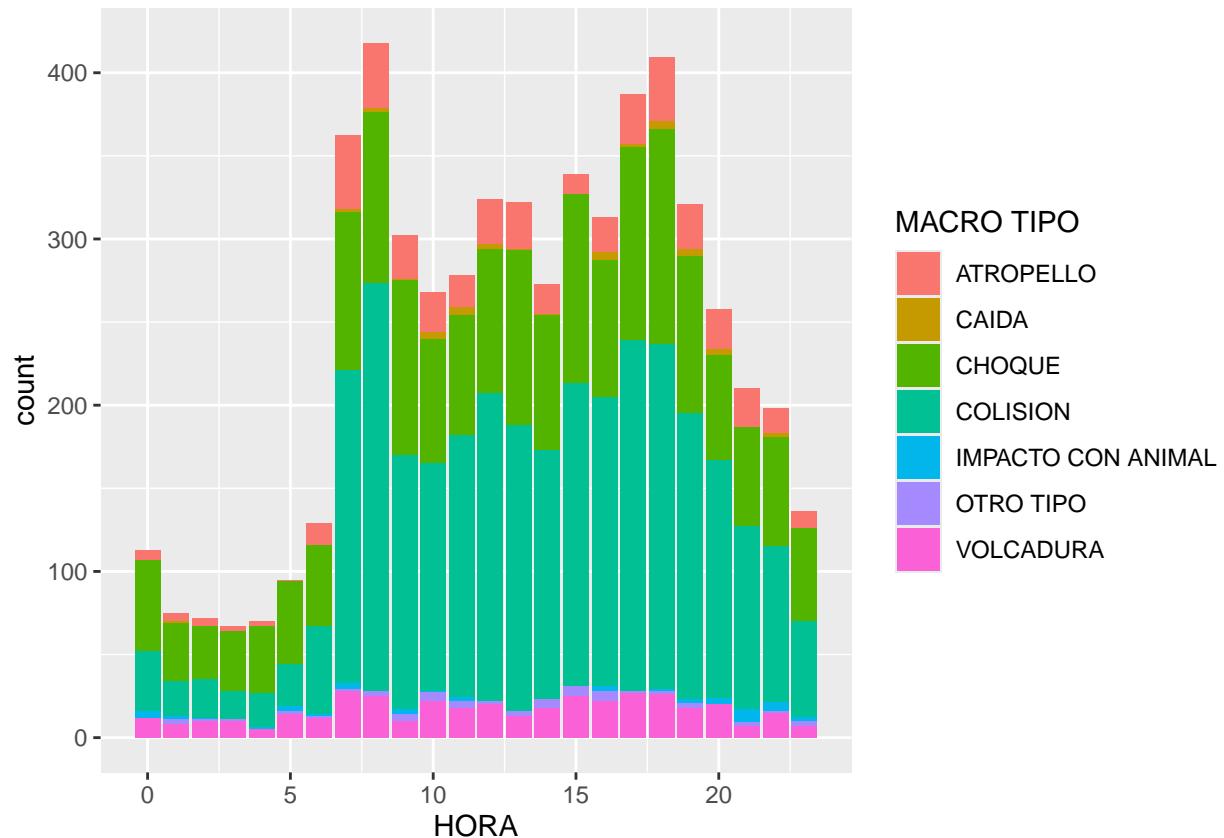
Distribución por fecha



Distribución por día de la semana

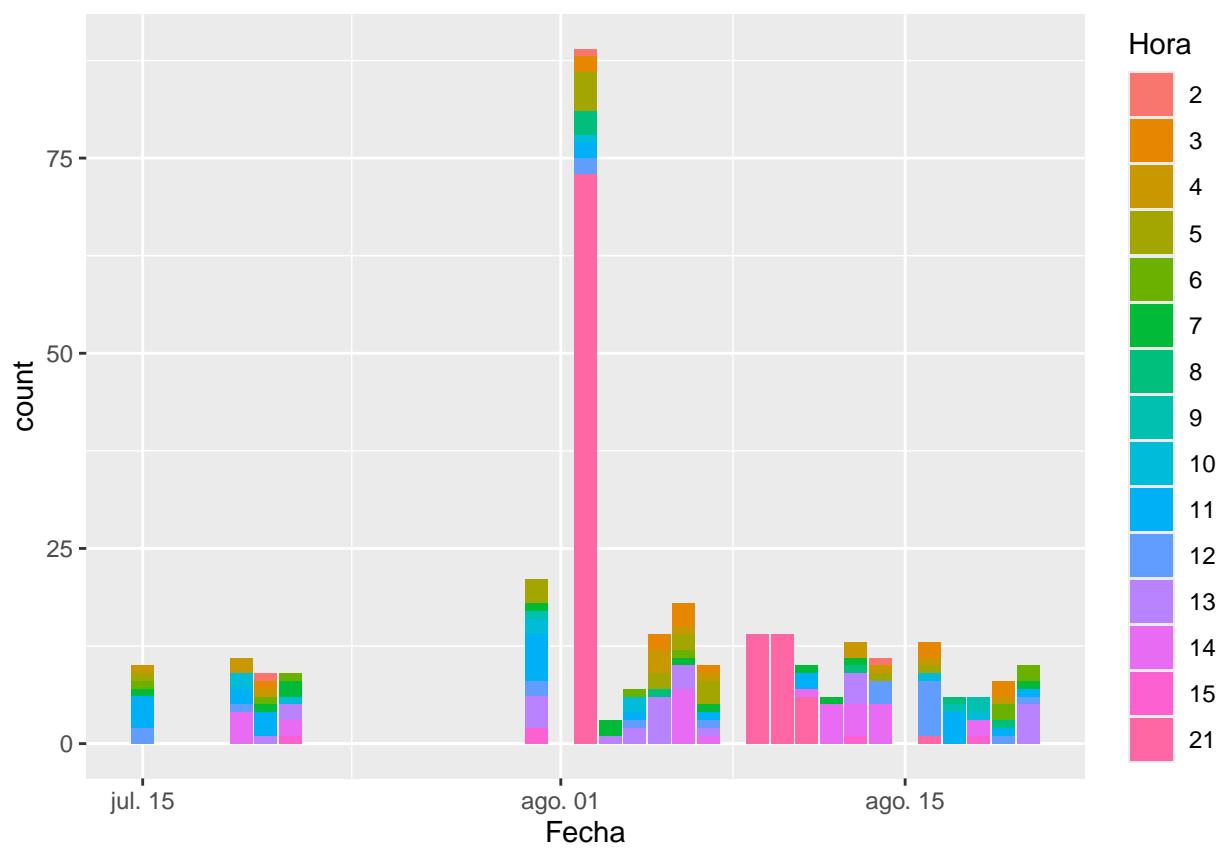


Distribución por hora

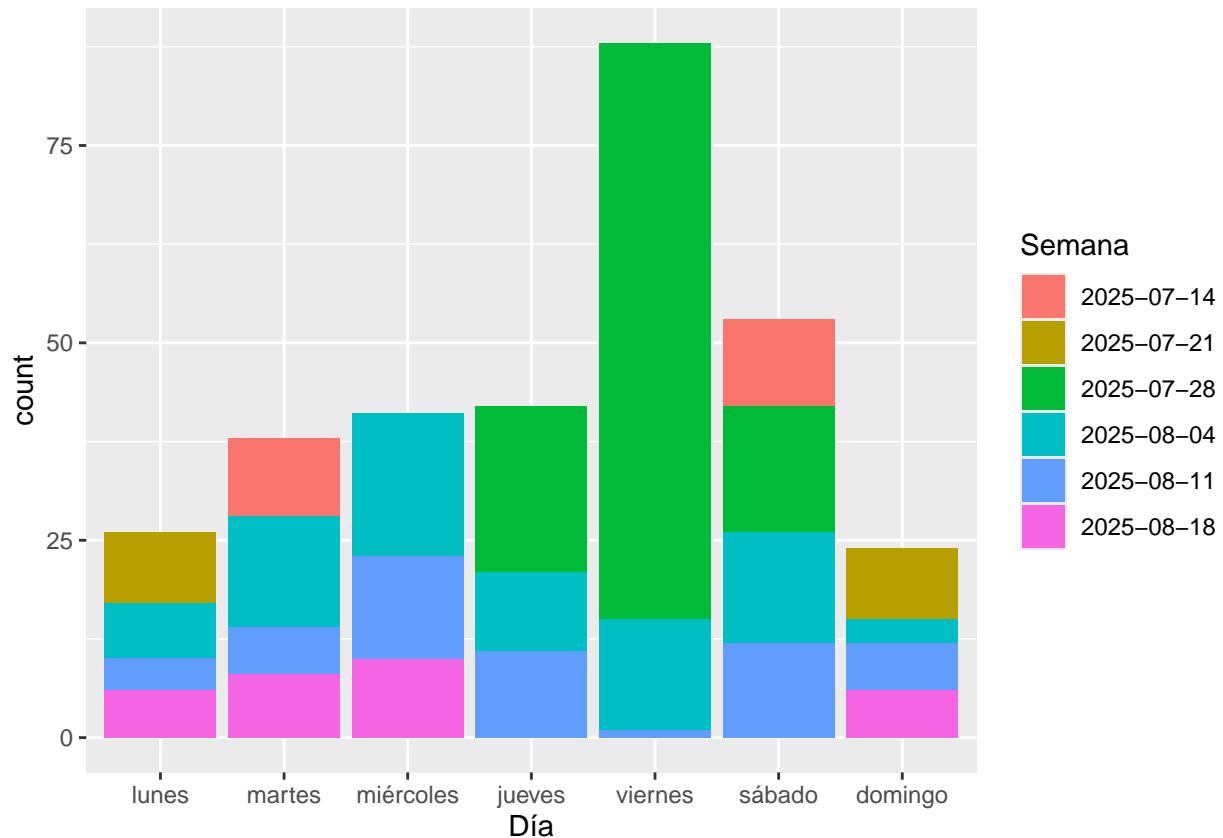


Incidentes reportados por radio

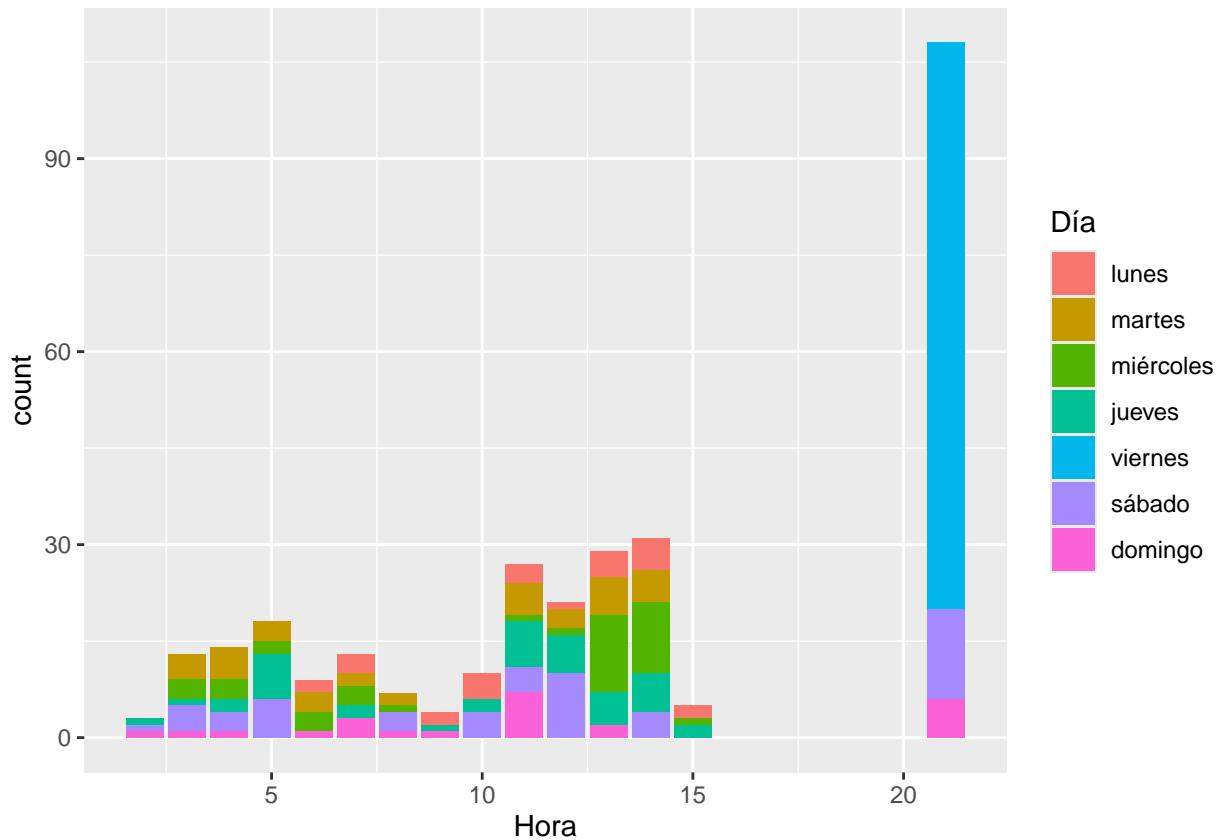
Distribución por fecha y hora



Distribución por día de la semana

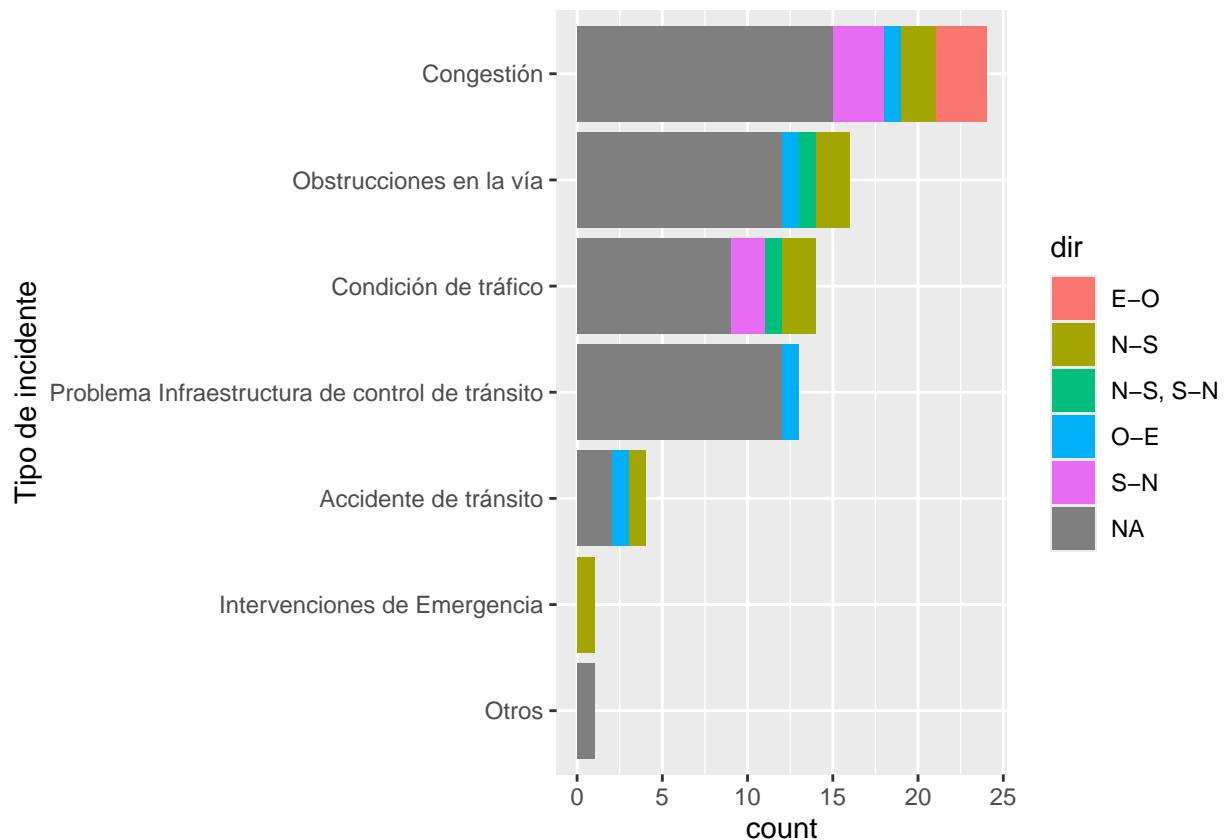


Distribución por hora



Hubo una gran cantidad de incidentes el día viernes 1ro de agosto a las 21 horas

Distribución por tipo de incidente durante el 1ro de agosto entre las 20 y 22 horas



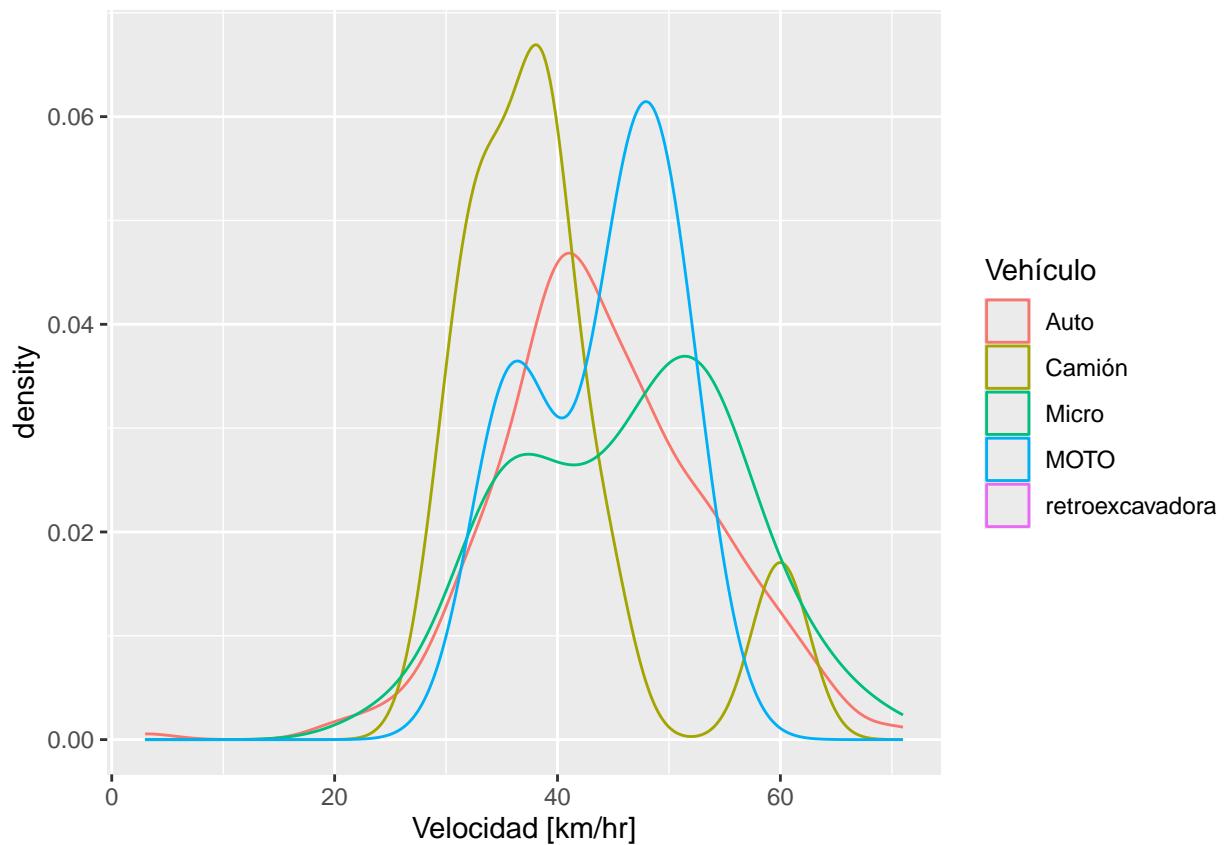
Distribución geográfica durante el 1ro de agosto entre las 20 y 22 horas

Zoom: 10

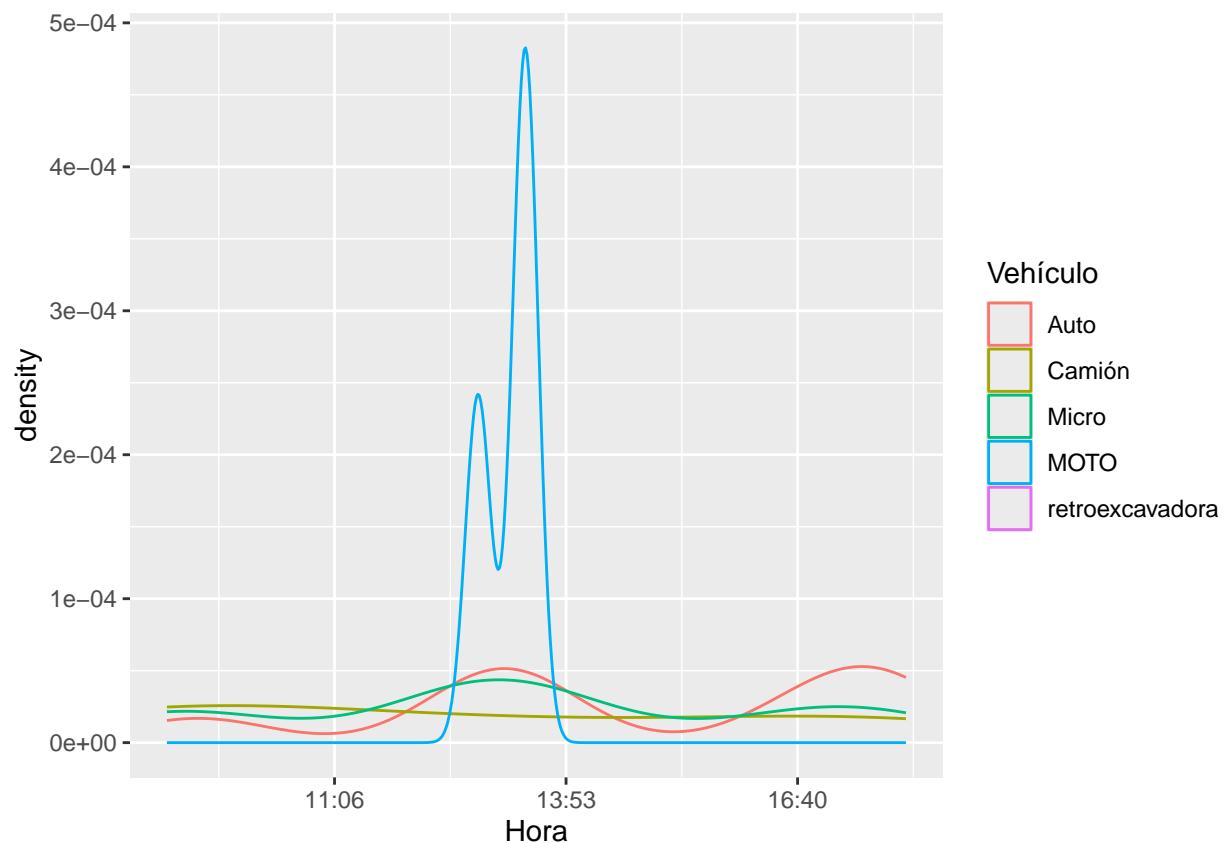


Mediciones de velocidad en paradero “Bdo O’Higgins - Arauco”, Chiguayante

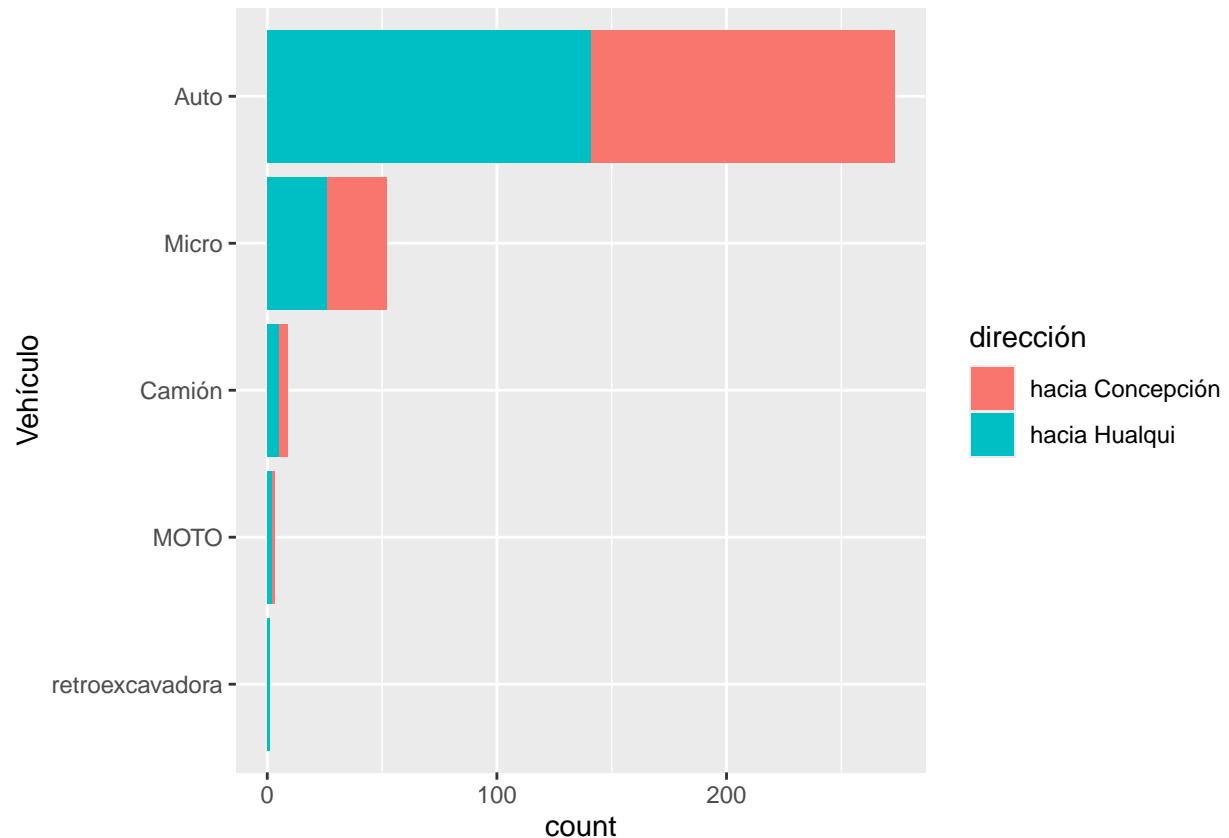
Velocidades por vehículo



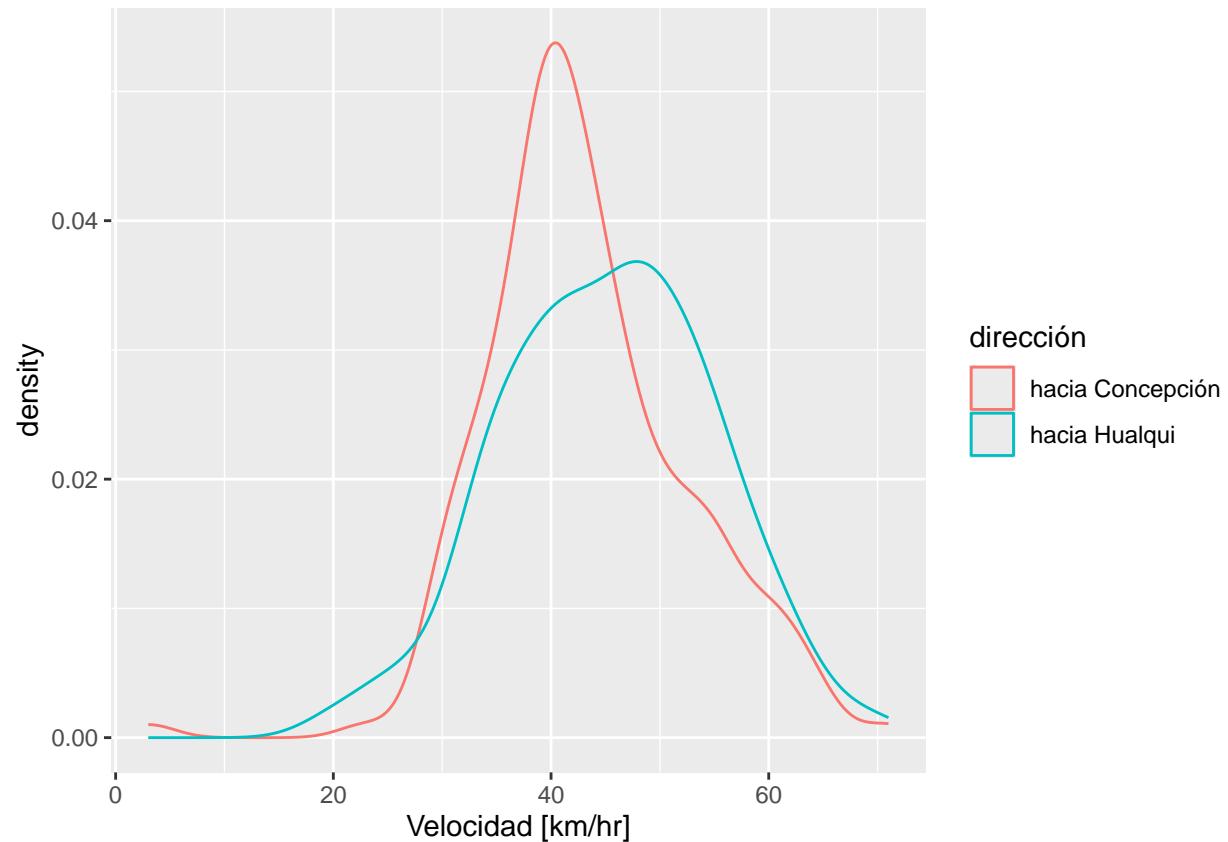
Horas por vehículo



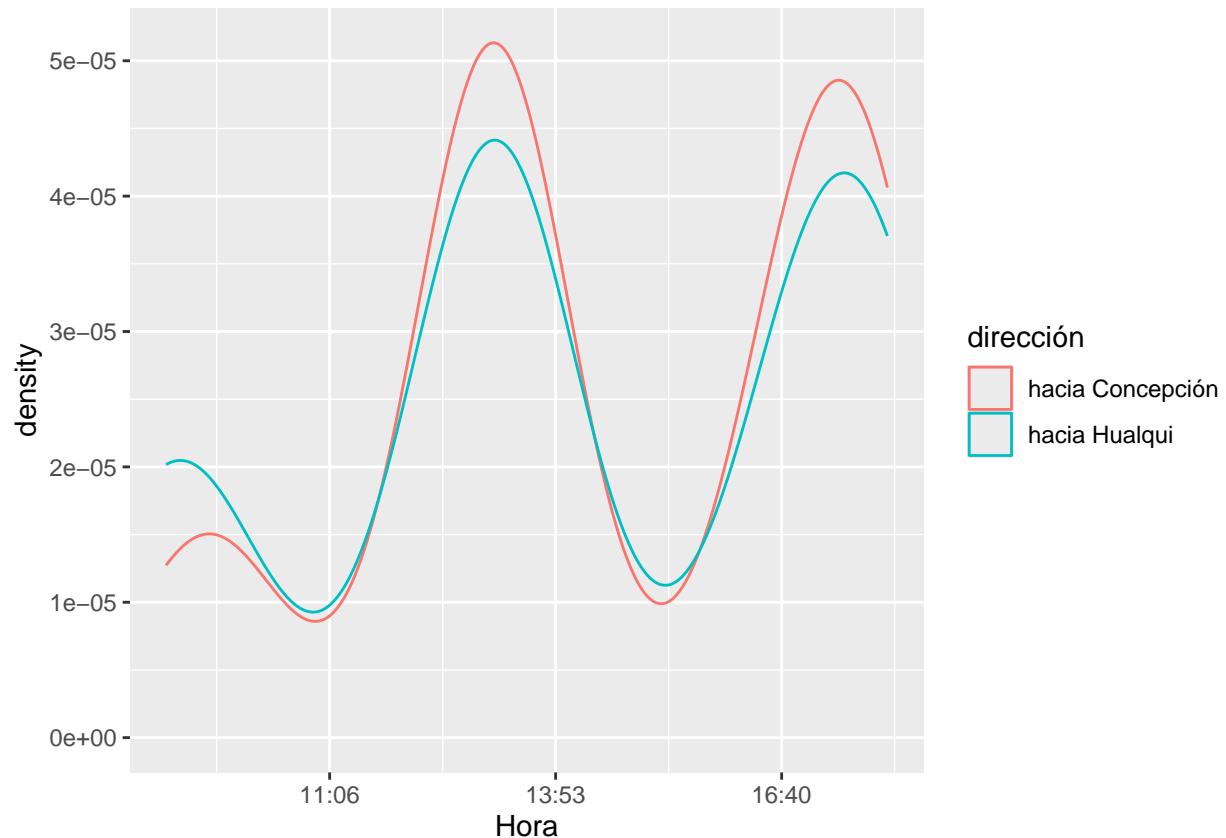
Distribución por vehículo



Velocidades por dirección

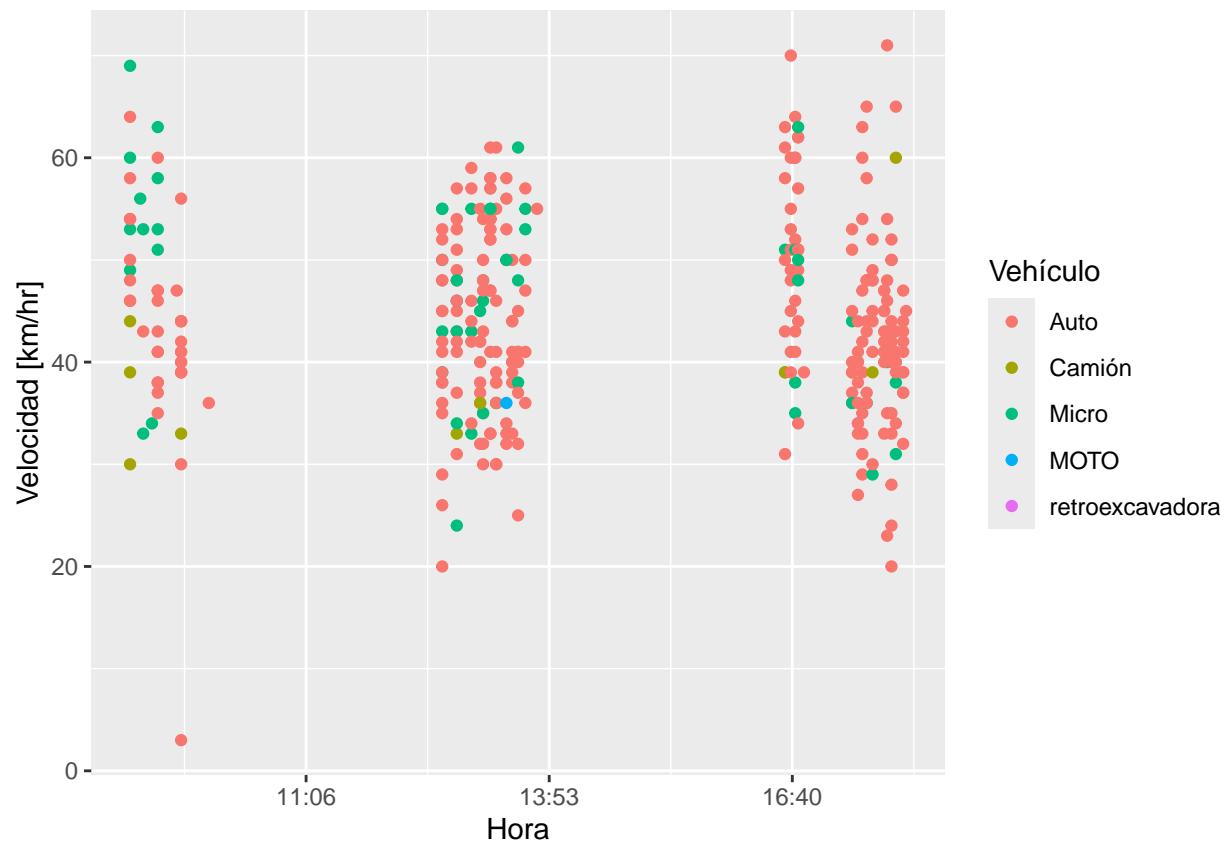


Horas por dirección

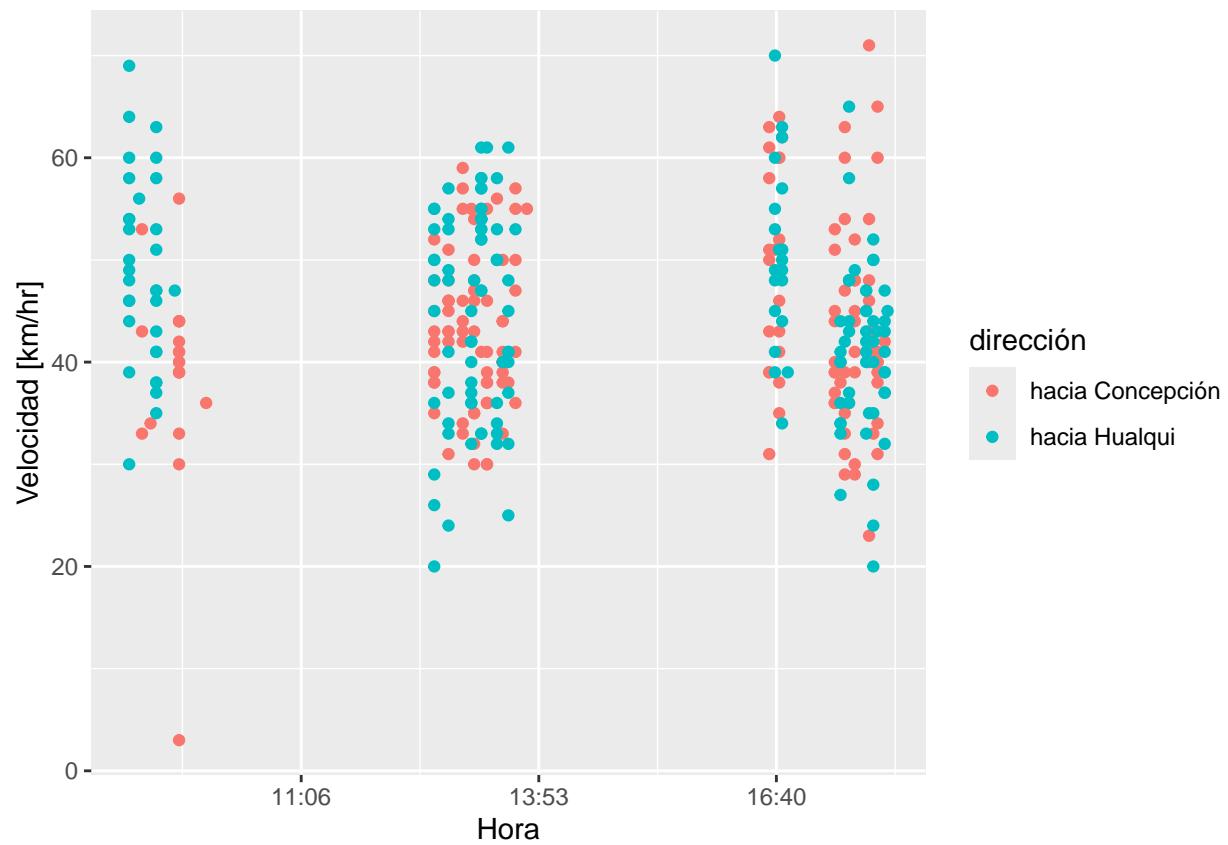


Distribución por Velocidad y Hora

Por Velocidad

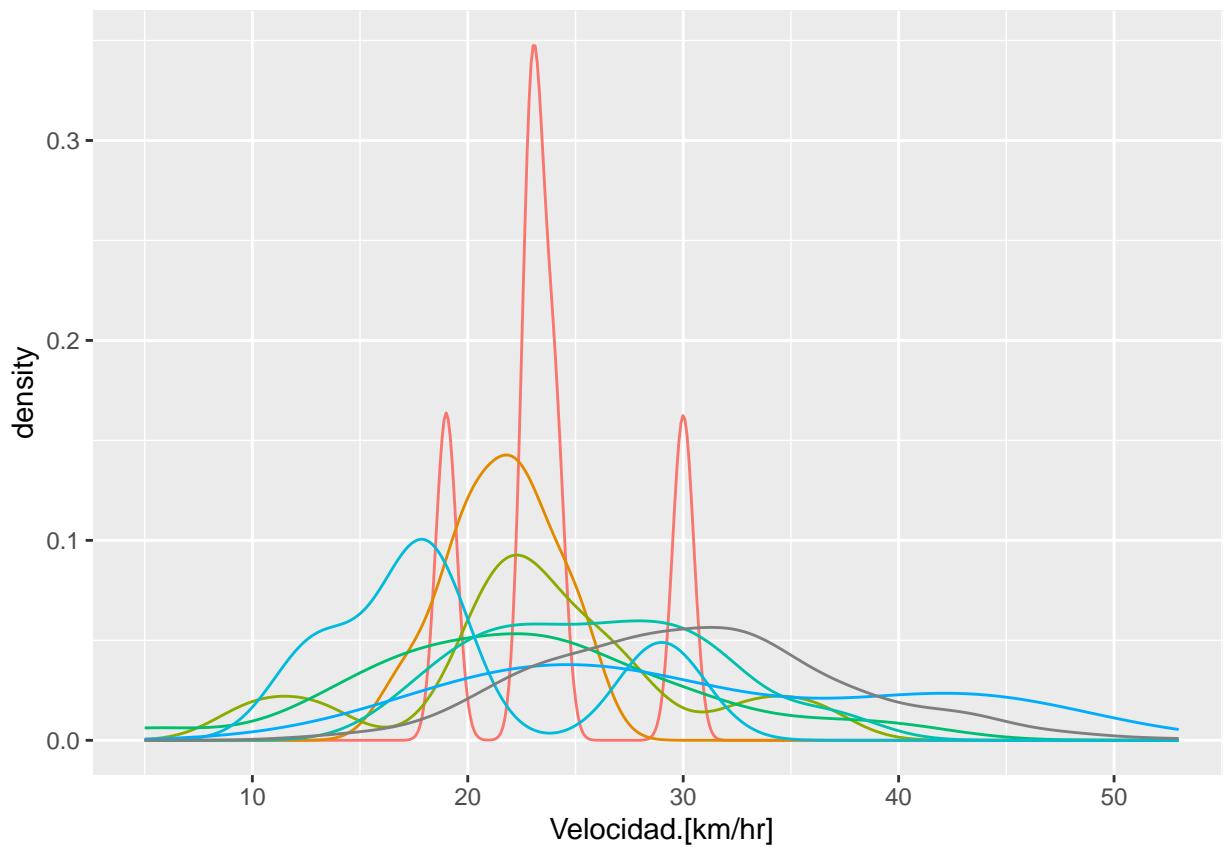


Por dirección



Mediciones de velocidad en el cruce de Briceños con Miraflores

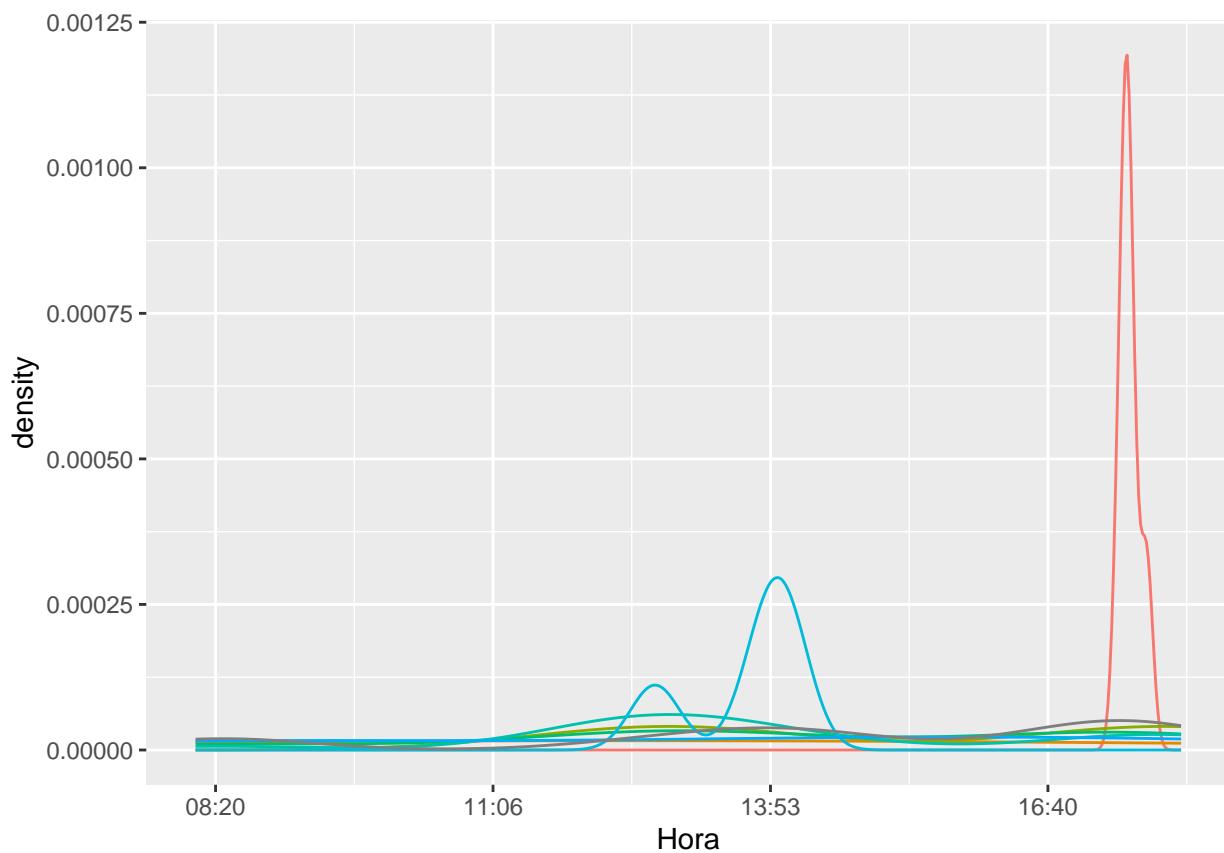
Velocidades por comentario



Comentario

- auto, llegando al reductor de velocidad
- camión de carga ubicado antes de reductor de velocidad (lomo de toro) autos con menor velocidad
- dobla en pasaje antes de lomo de toro
- Forman cola
- lejos, antes de lomo de toro
- Llega a cola
- Llegan juntos
- Micro
- Moto
- para antes de reductor de velocidad
- pasando el reductor de velocidad
- salta disco pare
- viene remolcando a otro, camiones
- NA

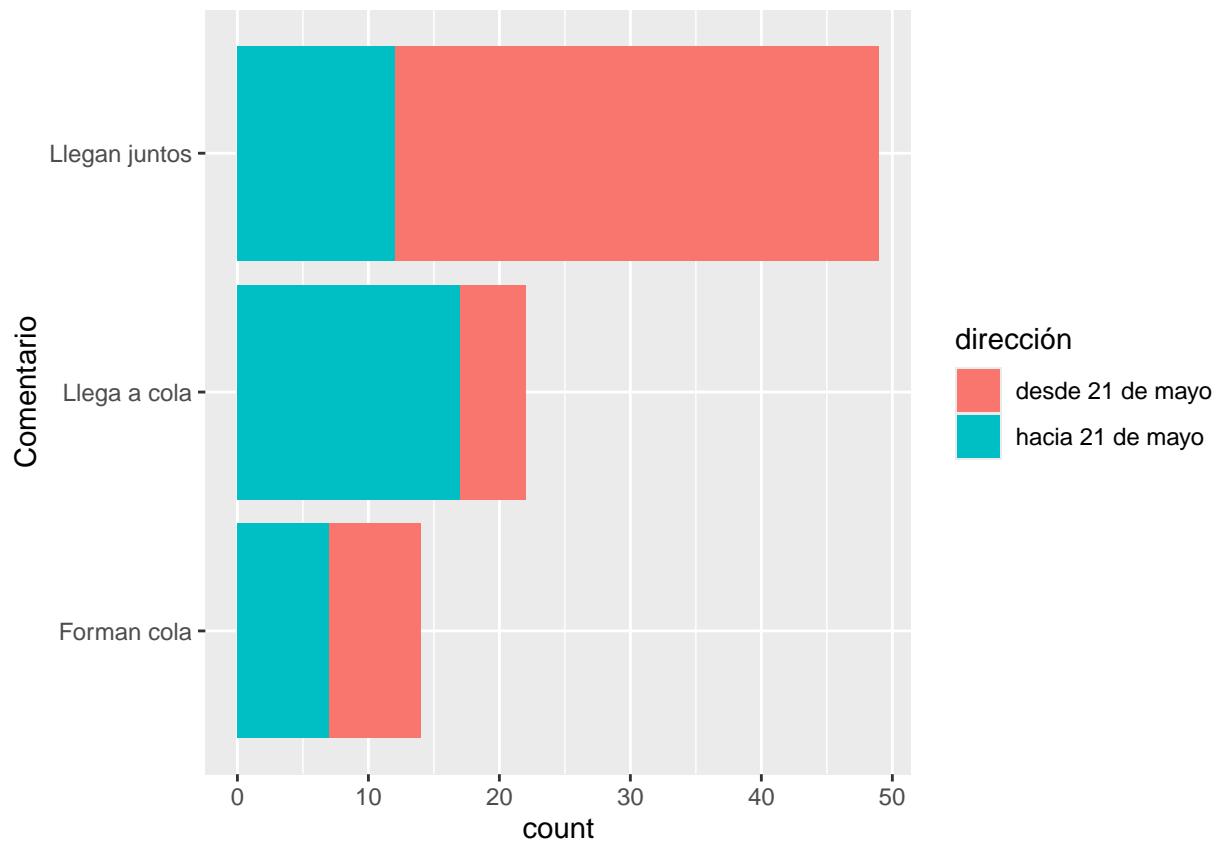
Horas por comentario



Comentario

- auto, llegando al reductor de velocidad
- camión de carga ubicado antes de reductor de velocidad (lomo de toro) autos con menor velocidad
- dobla en pasaje antes de lomo de toro
- Forman cola
- lejos, antes de lomo de toro
- Llega a cola
- Llegan juntos
- Micro
- Moto
- para antes de reductor de velocidad
- pasando el reductor de velocidad
- salta disco pare
- viene remolcando a otro, camiones
- NA

3 comentarios más frecuentes

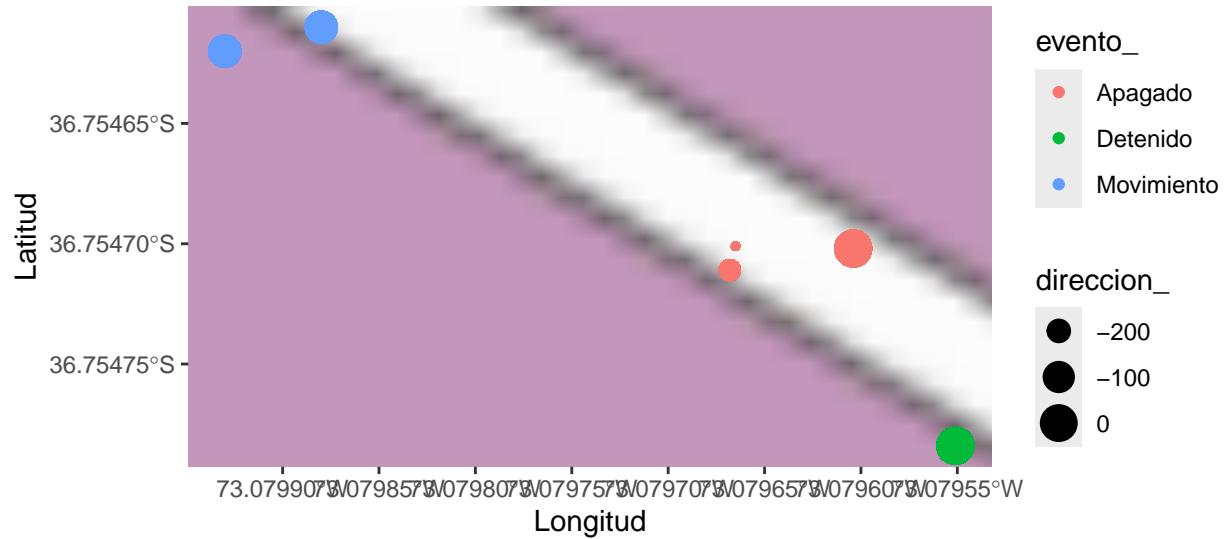


Alertas GPS

Eventos por localización

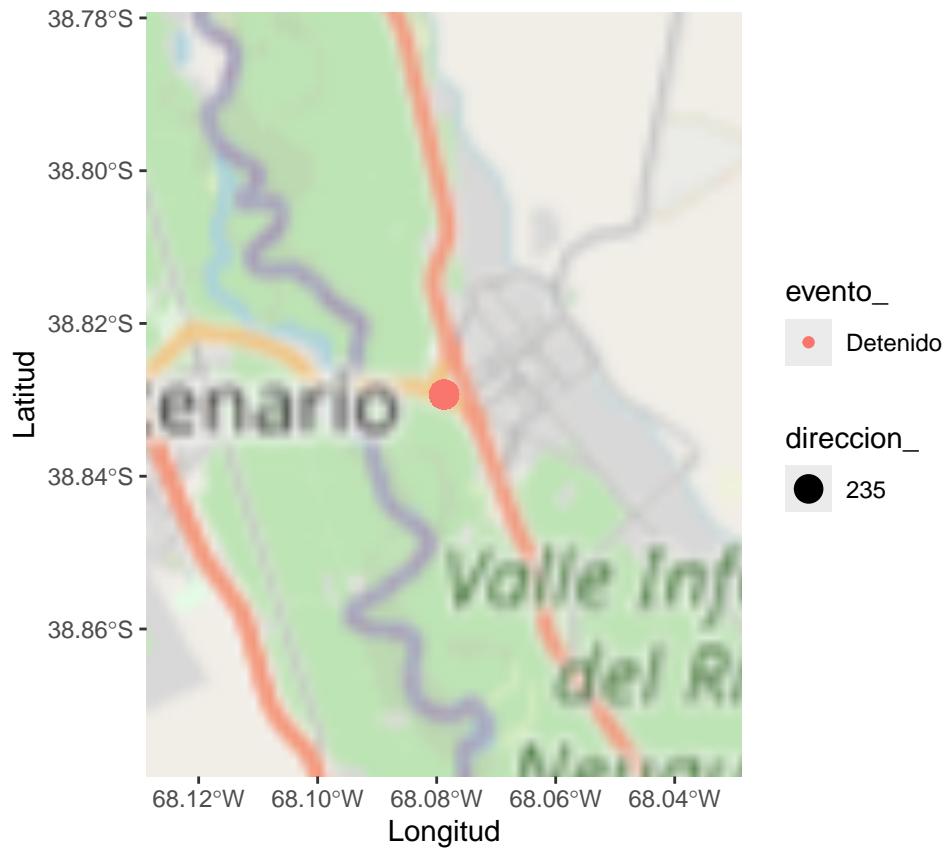
Chile

Zoom: 17



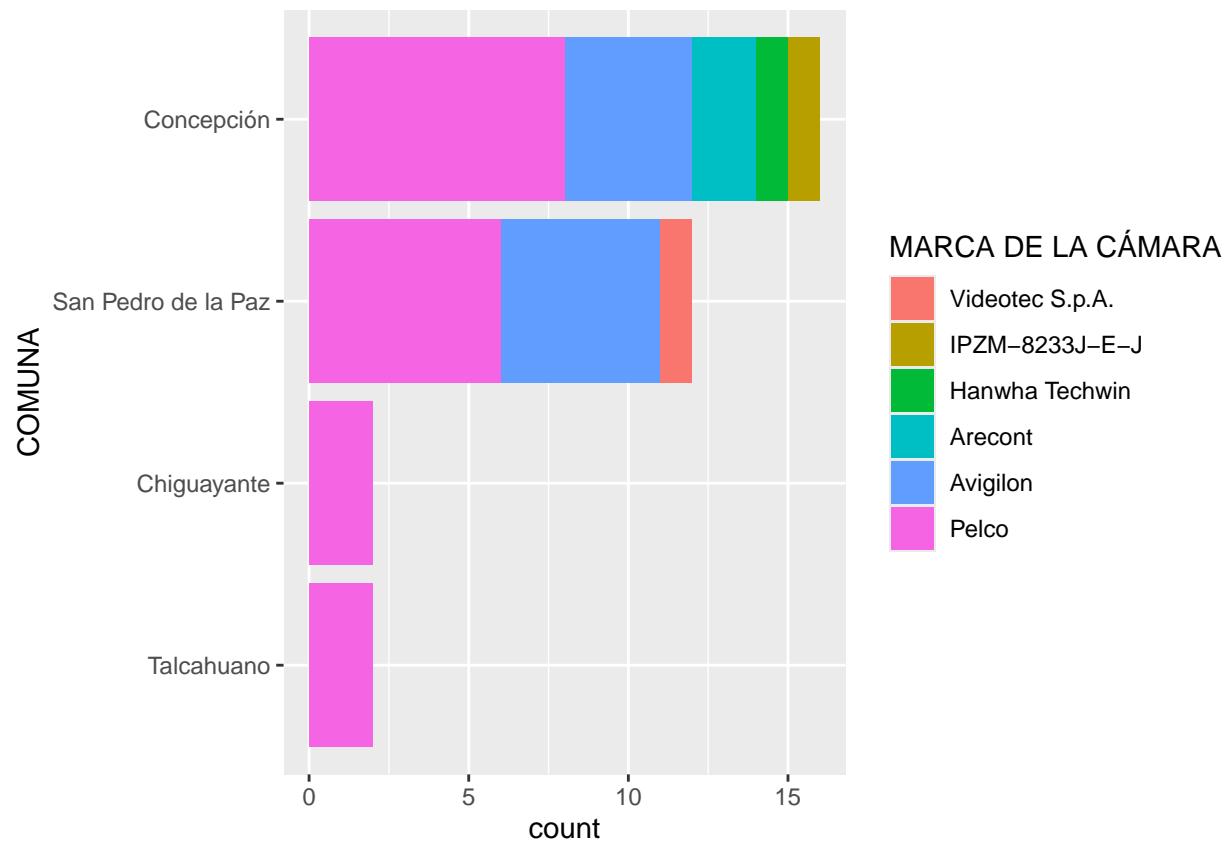
Argentina

Zoom: 10

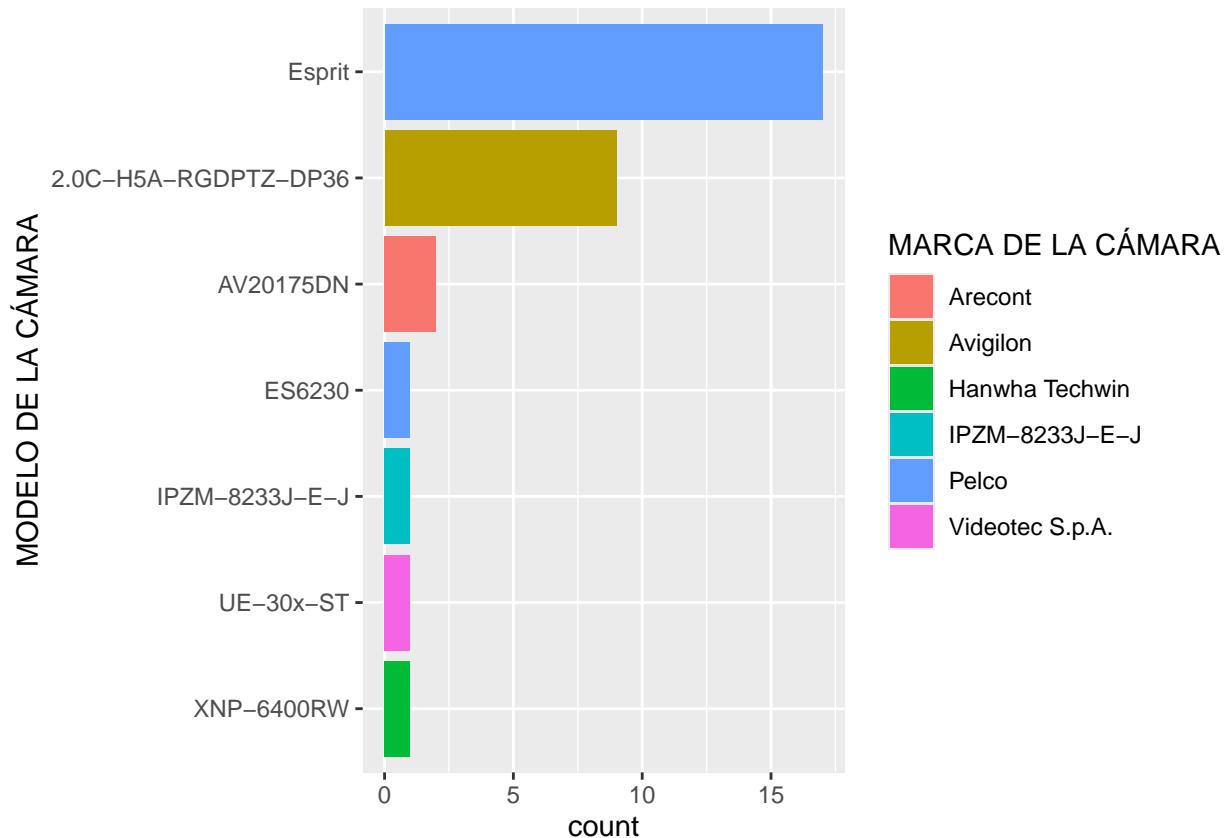


Inventario de de cámaras de vigilancia

Distribución por columna



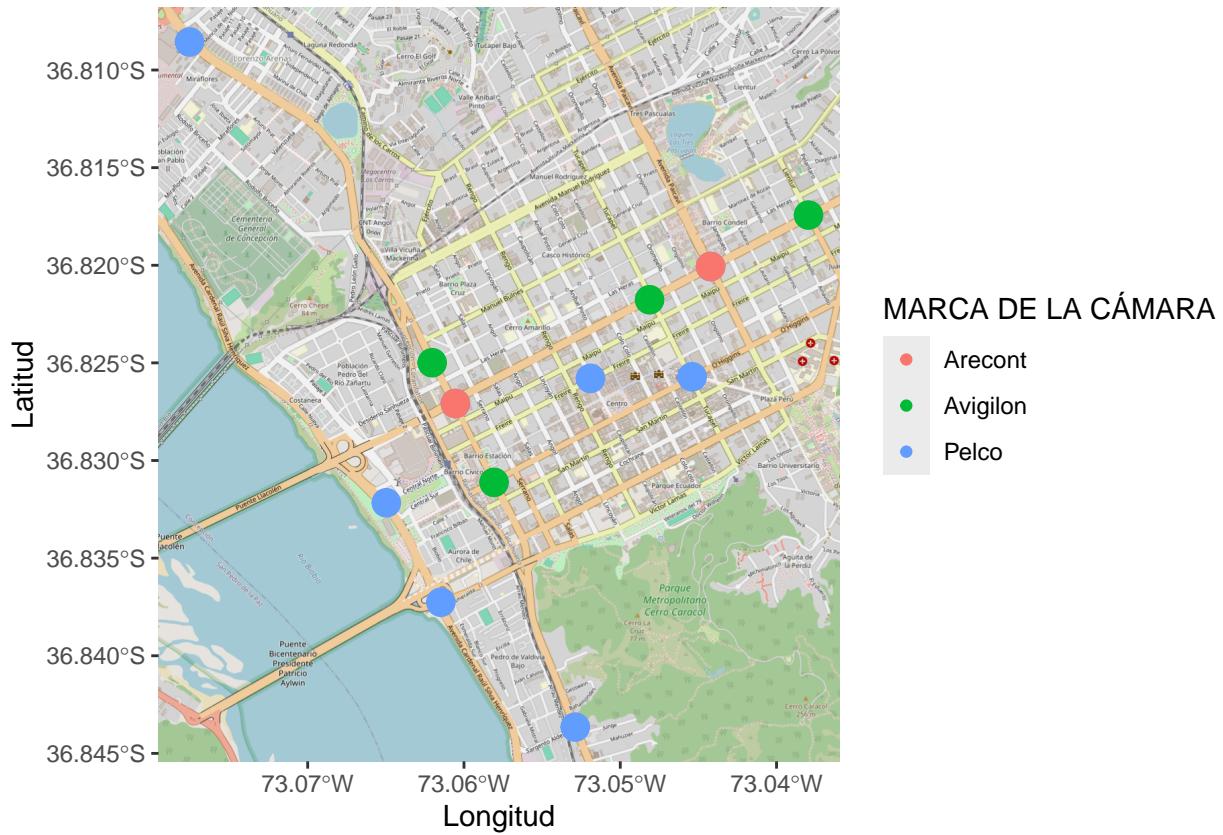
Distribución por modelo



Marca de la cámara por localización

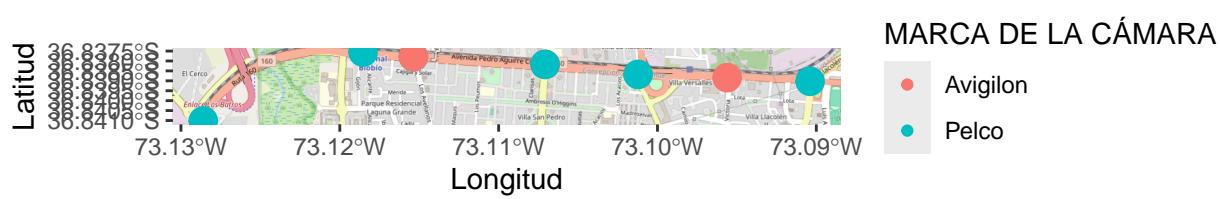
Concepción

Zoom: 15



San Pedro de la Paz

Zoom: 15



Chiguayante

Zoom: 15

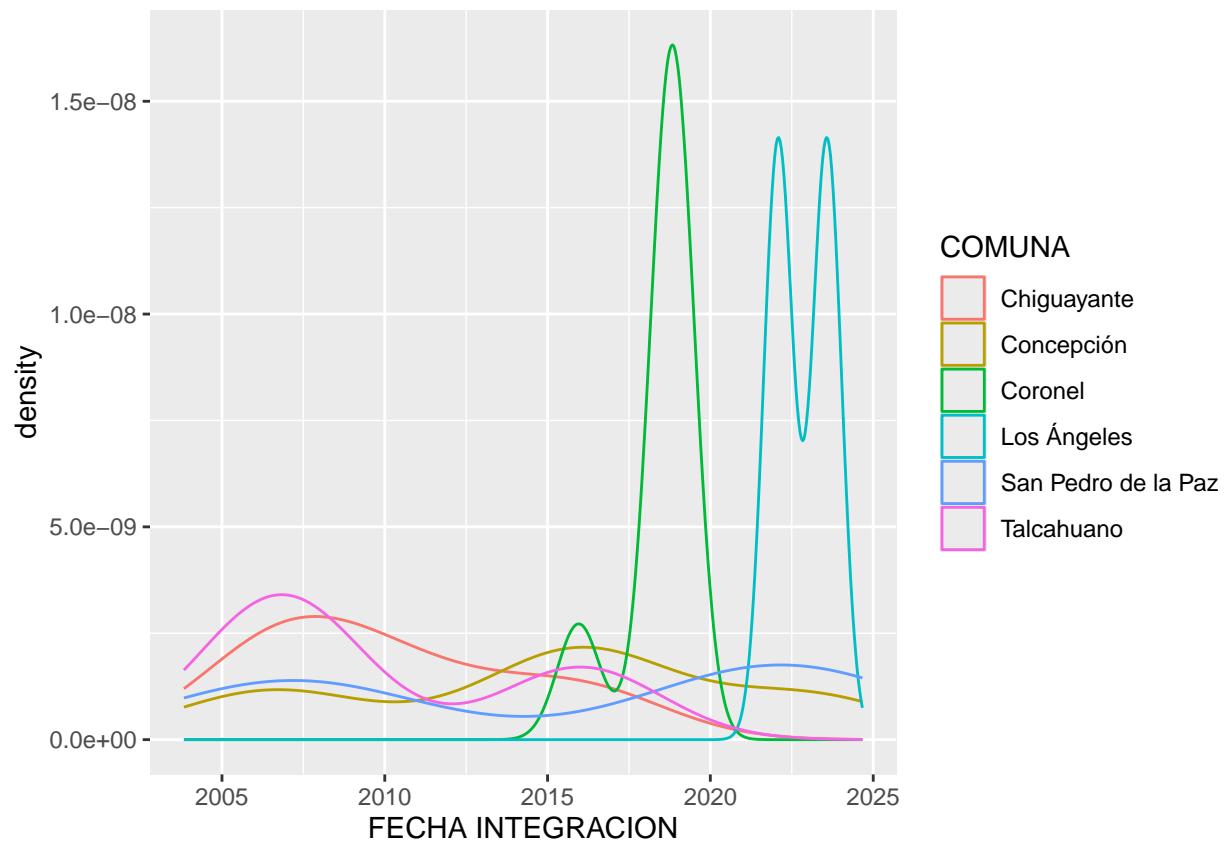


Talcahuano

Zoom: 19

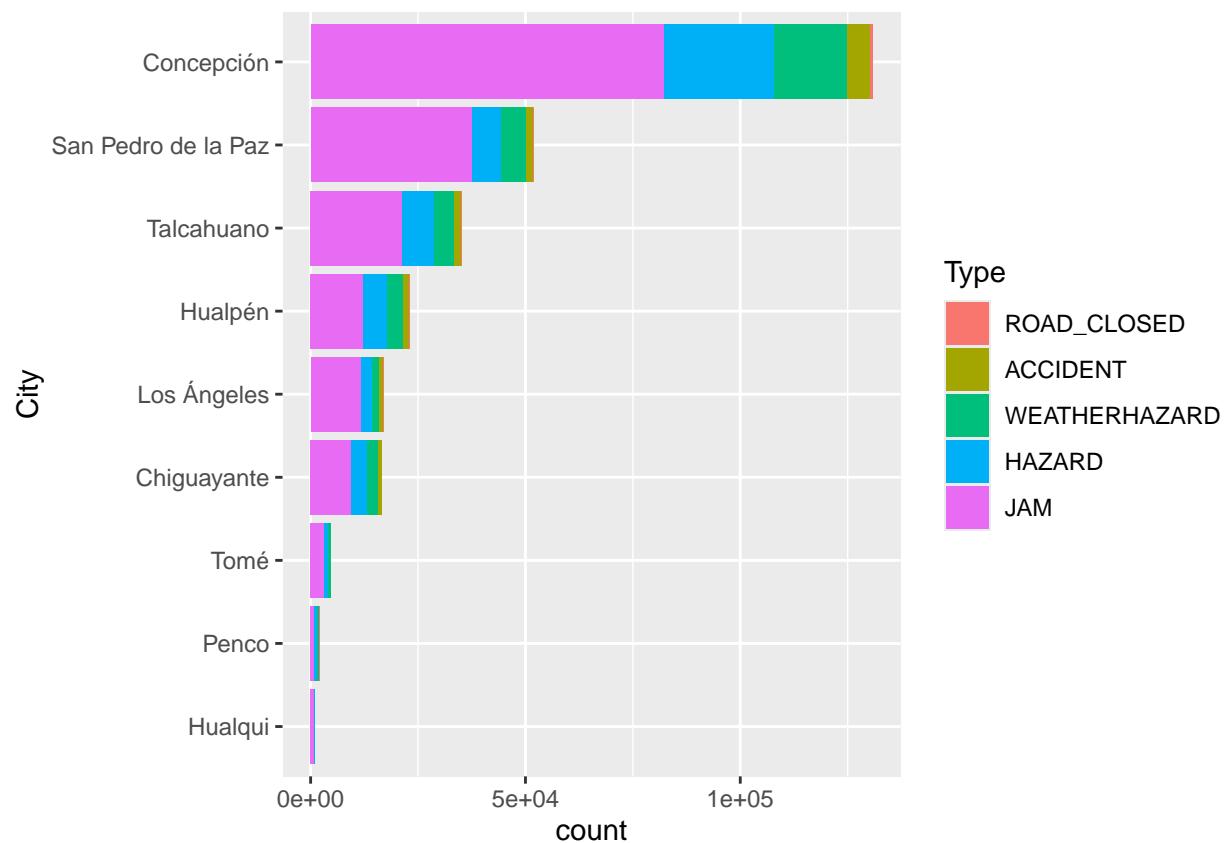


Distribución de las fechas de integración

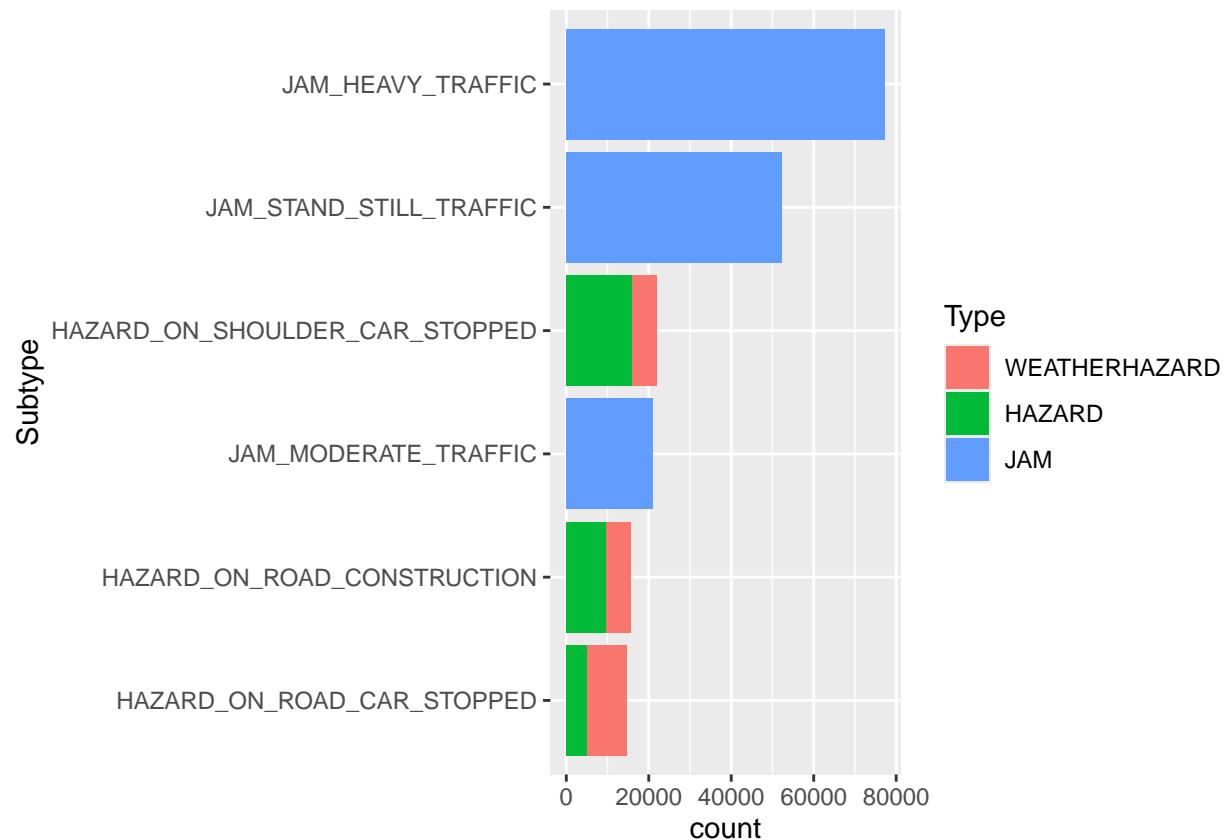


Alertas de Tráfico

Distribución por ciudad



Distribución por subtipo

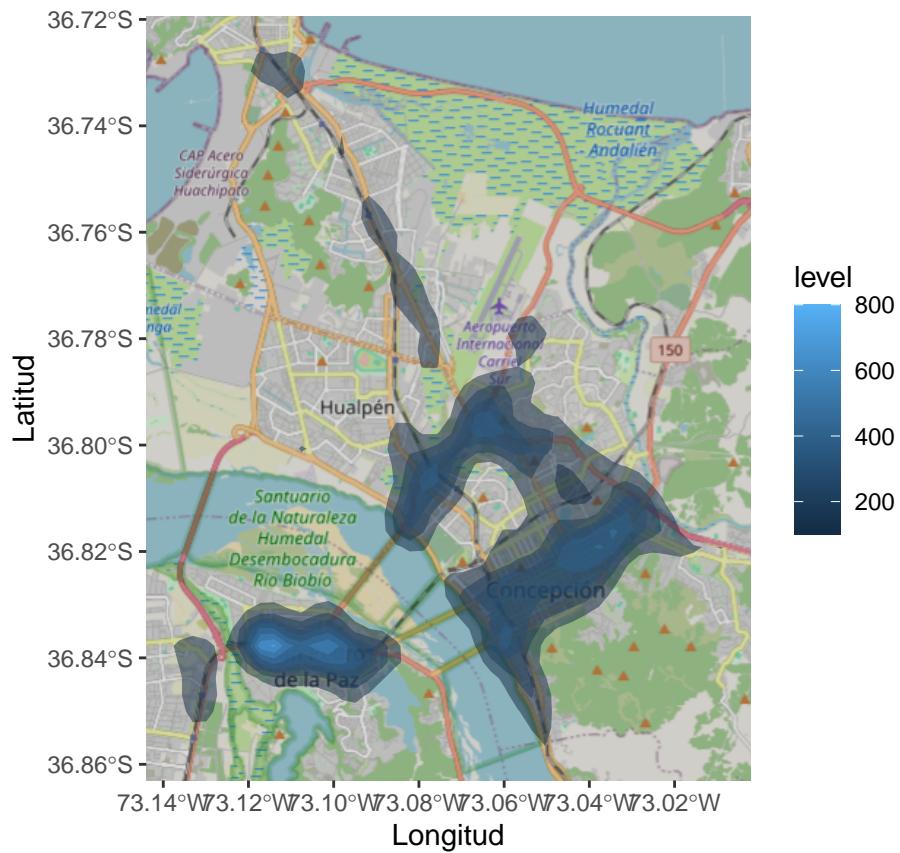


Distribución geográfica

Concepción, San Pedro de la Paz, Talcahuano, Hualpén, Penco y Hualqui

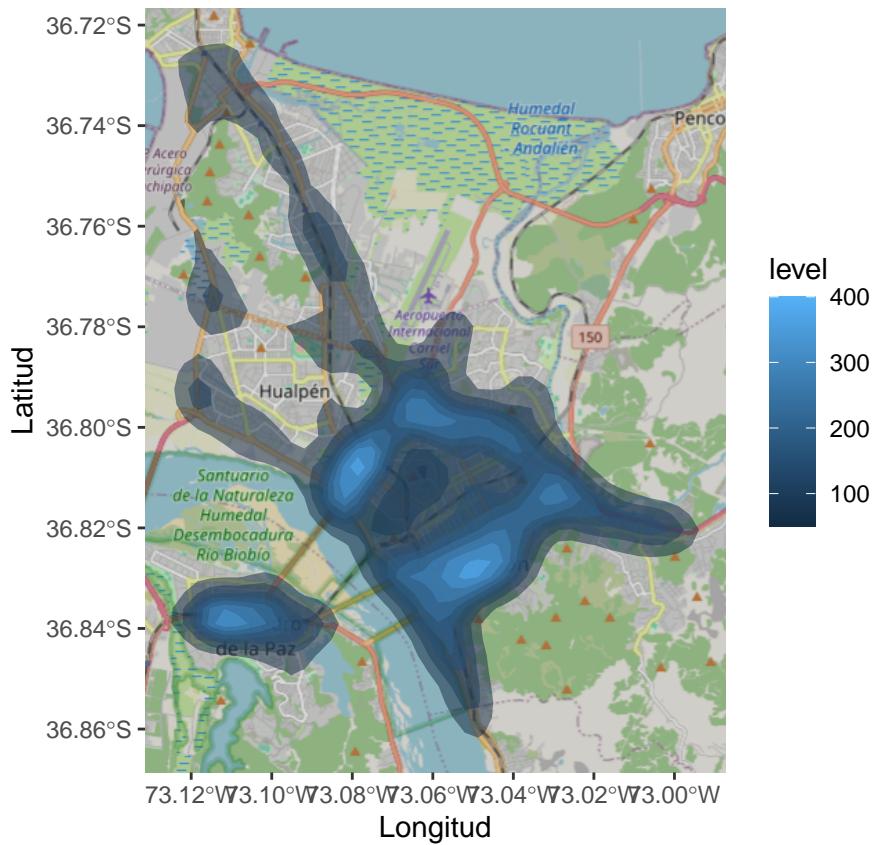
JAM

Zoom: 12



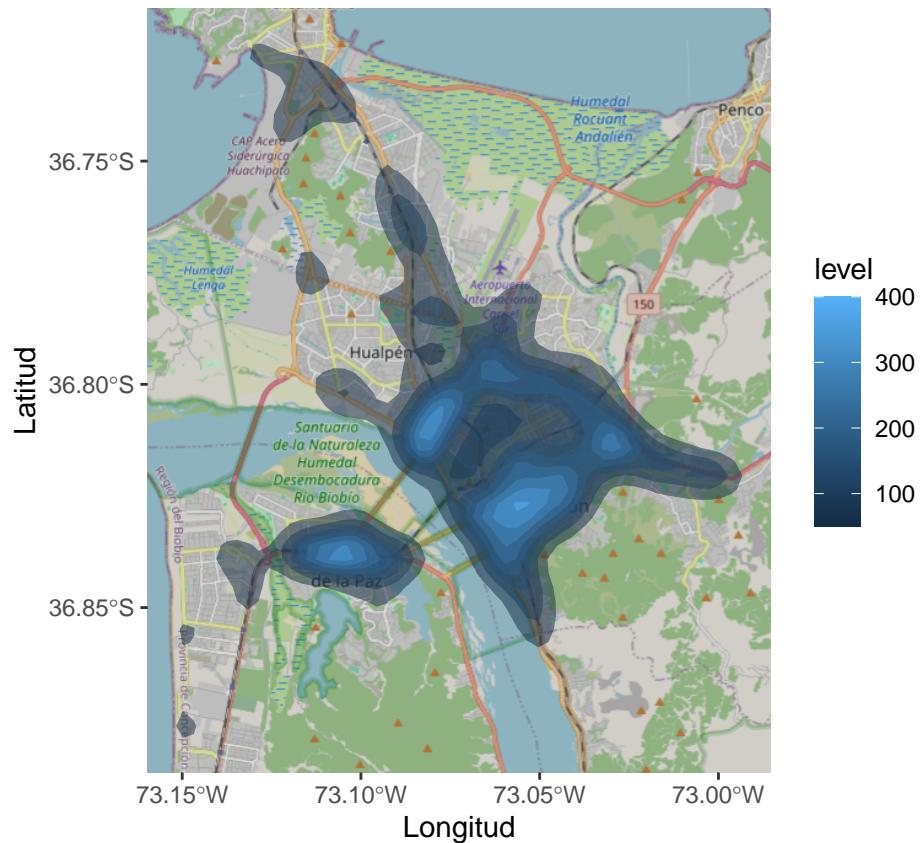
HAZARD

Zoom: 12



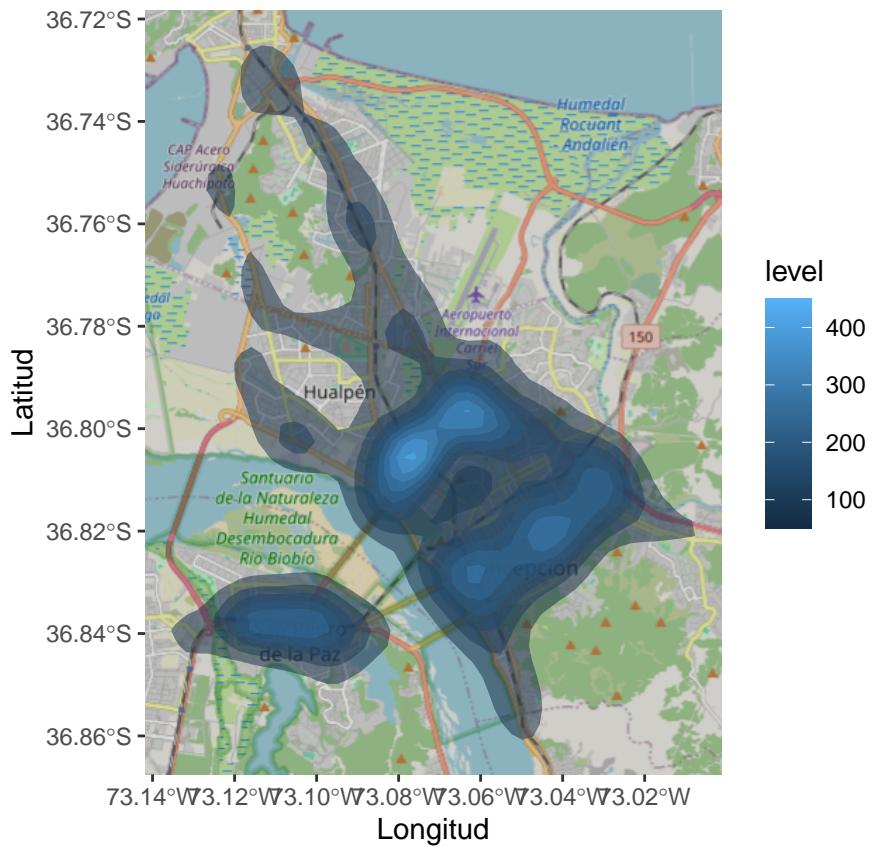
WEATHERHAZARD

Zoom: 12



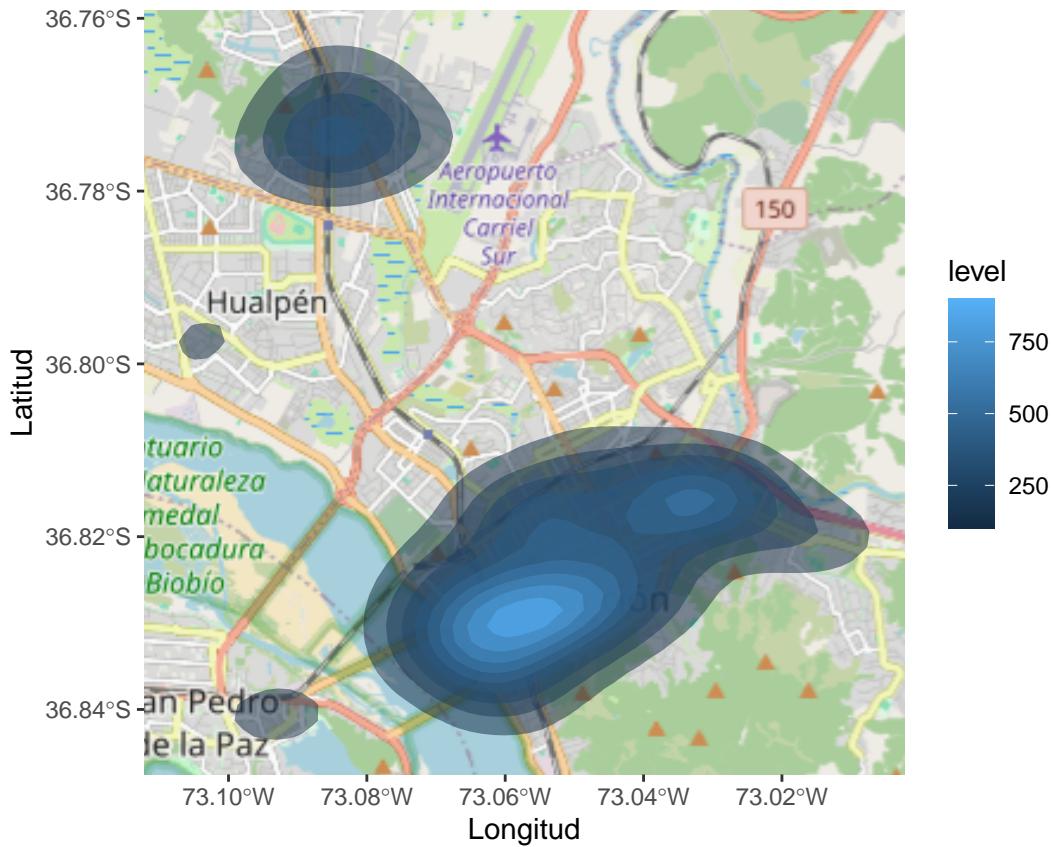
ACCIDENT

Zoom: 12



ROAD_CLOSED

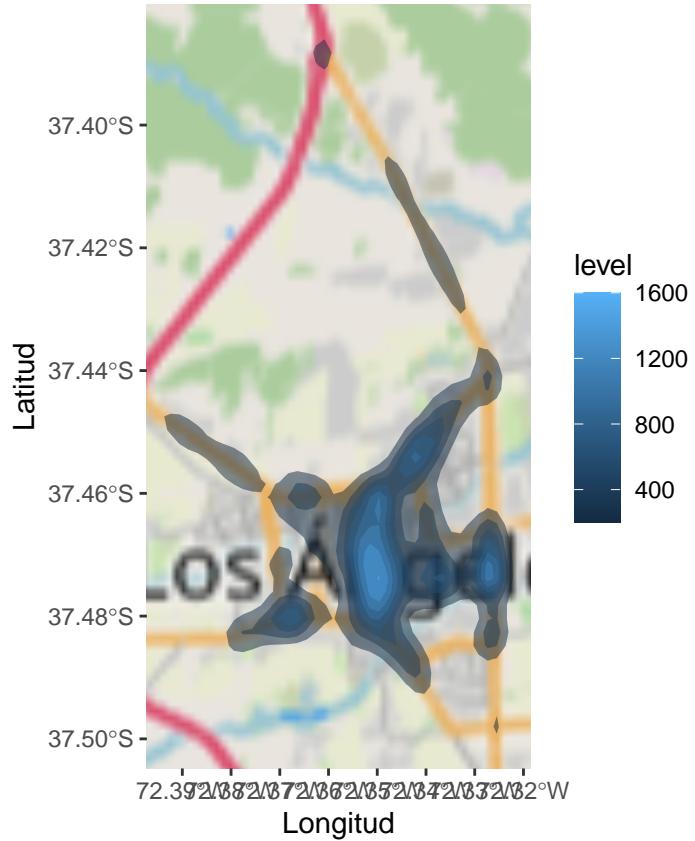
Zoom: 12



Los Ángeles

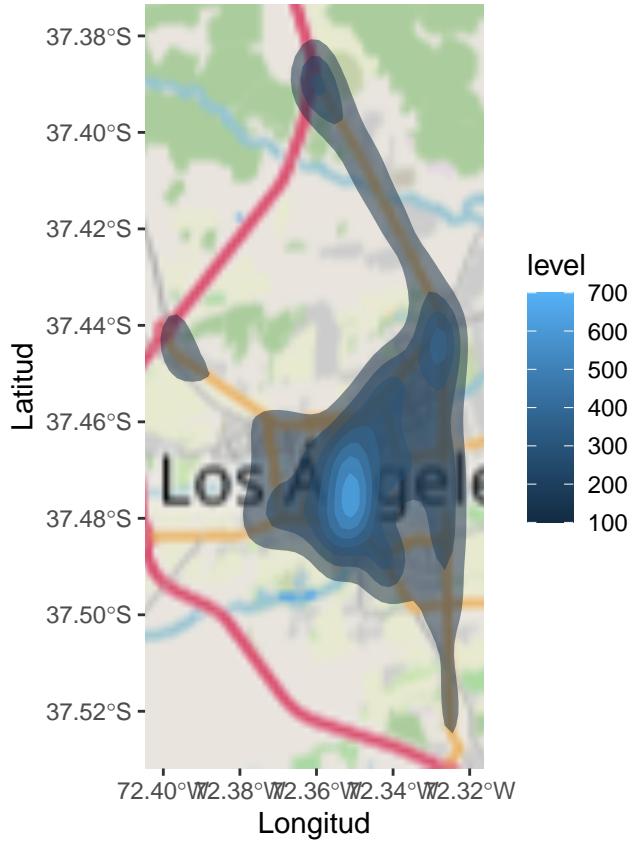
JAM

Zoom: 10



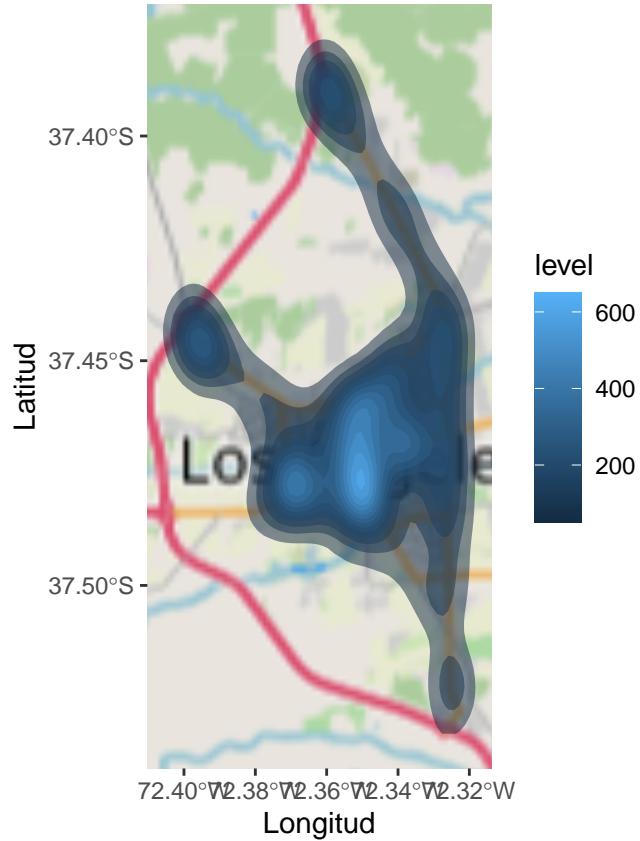
HAZARD

Zoom: 10



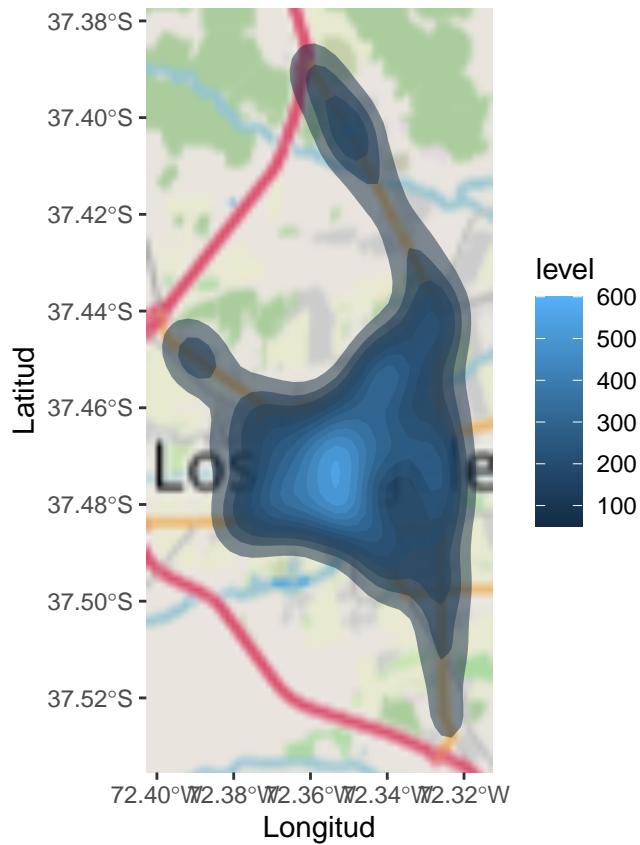
WEATHERHAZARD

Zoom: 10



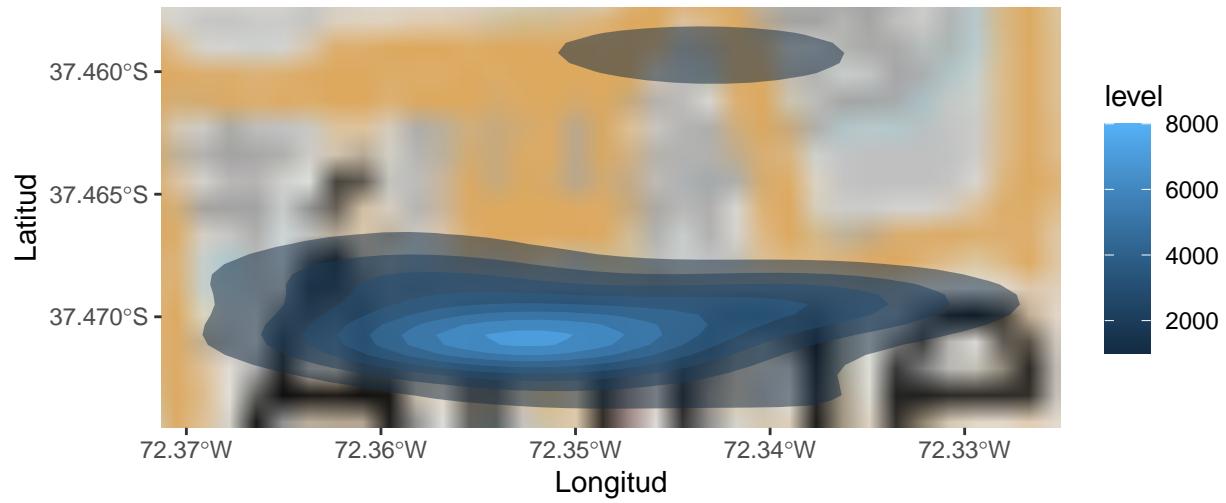
ACCIDENT

Zoom: 10



ROAD_CLOSED

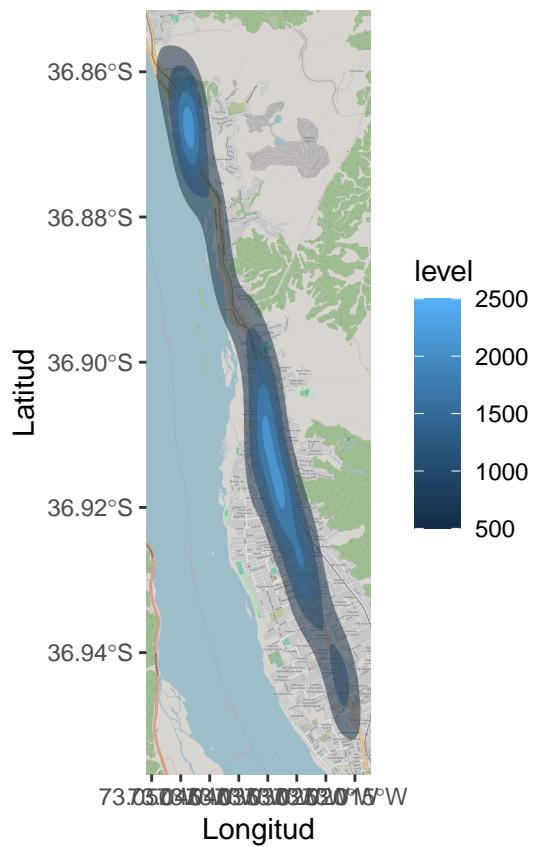
Zoom: 10



Chiguayante

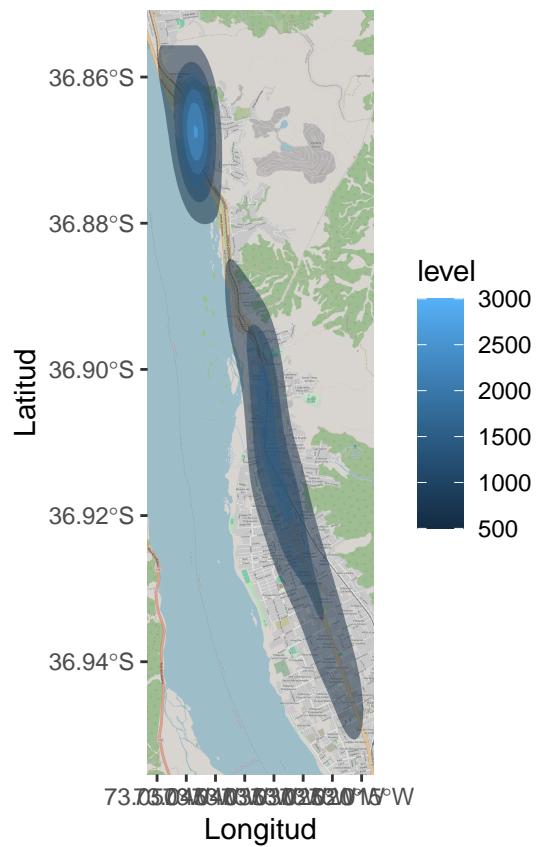
JAM

Zoom: 15



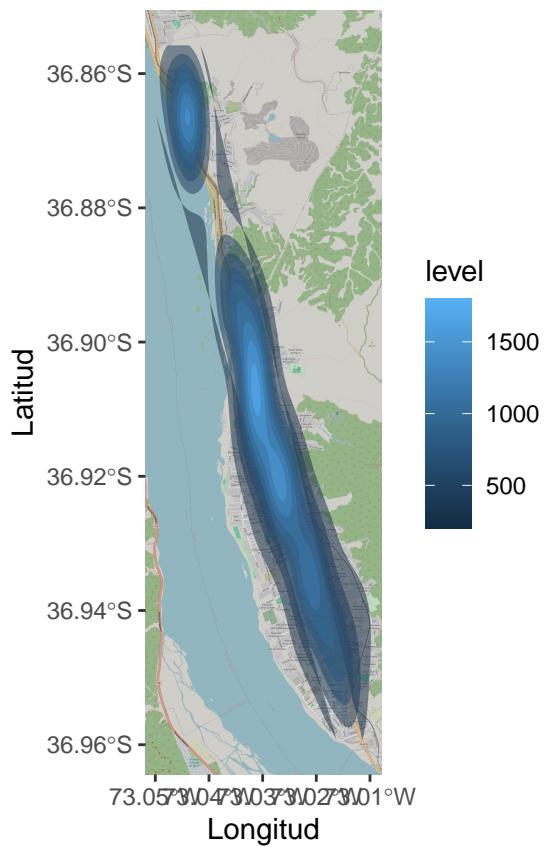
HAZARD

Zoom: 15



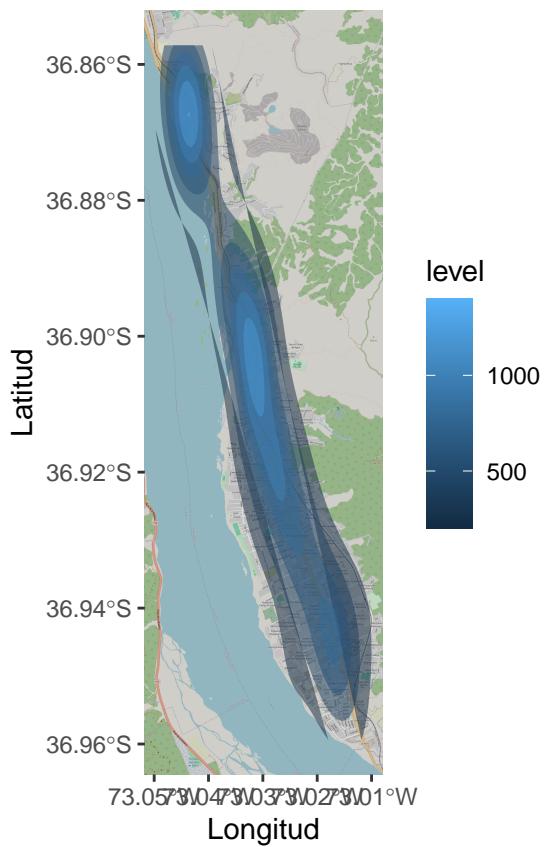
WEATHERHAZARD

Zoom: 15



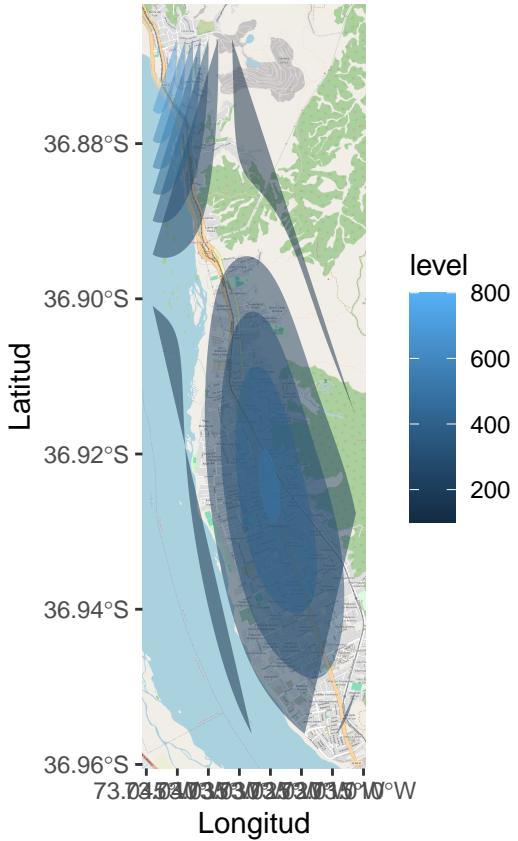
ACCIDENT

Zoom: 15



ROAD_CLOSED

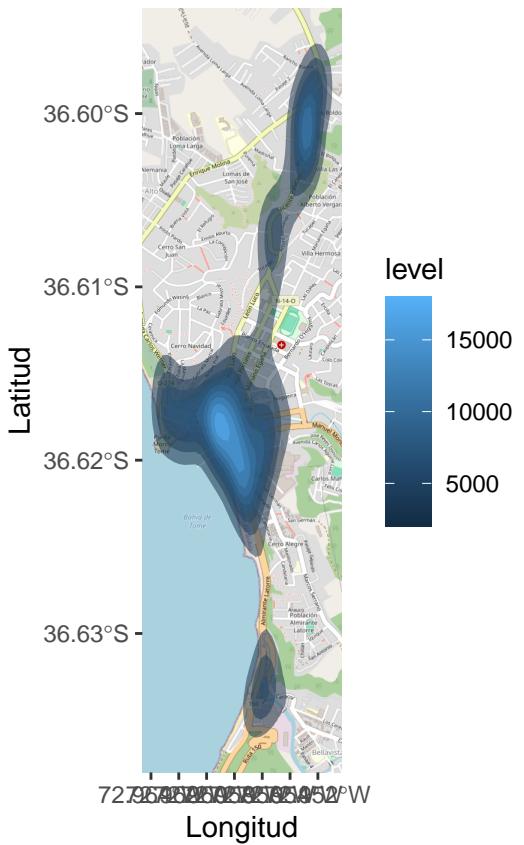
Zoom: 15



Tomé

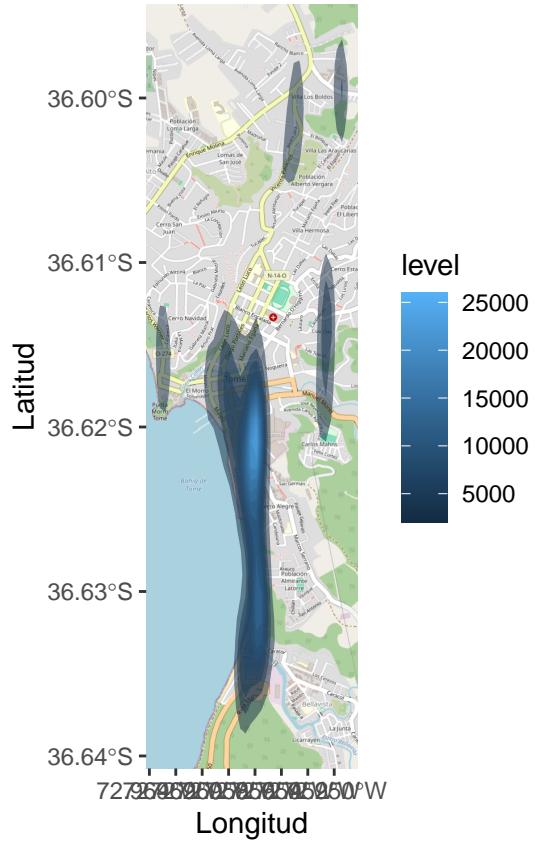
JAM

Zoom: 15



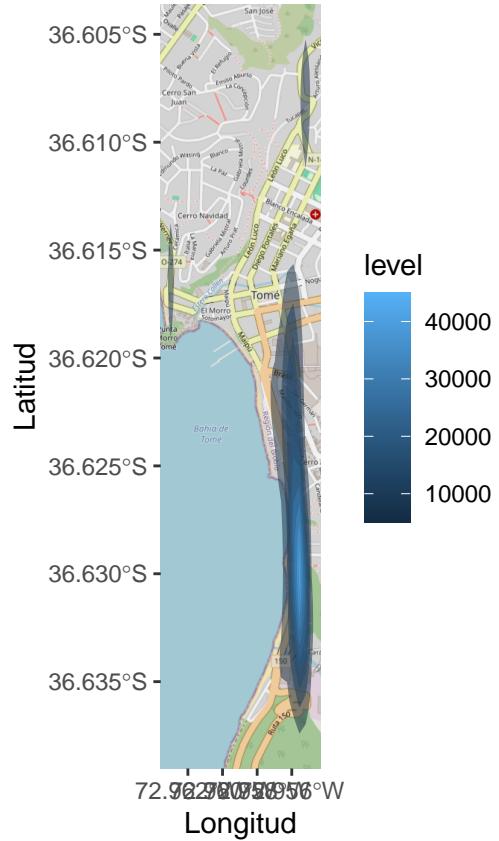
HAZARD

Zoom: 15



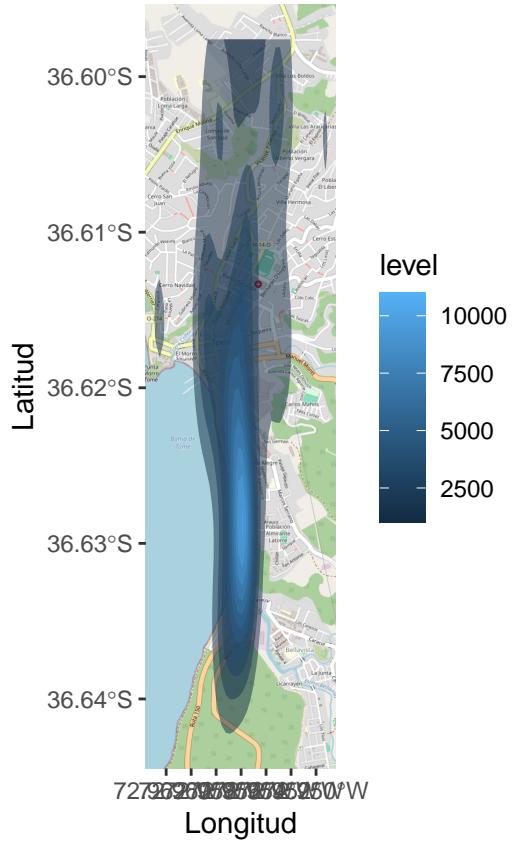
WEATHERHAZARD

Zoom: 15



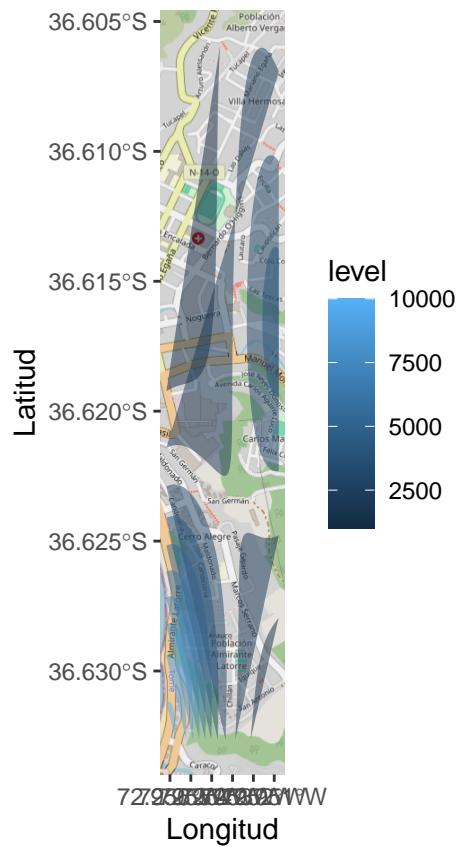
ACCIDENT

Zoom: 15



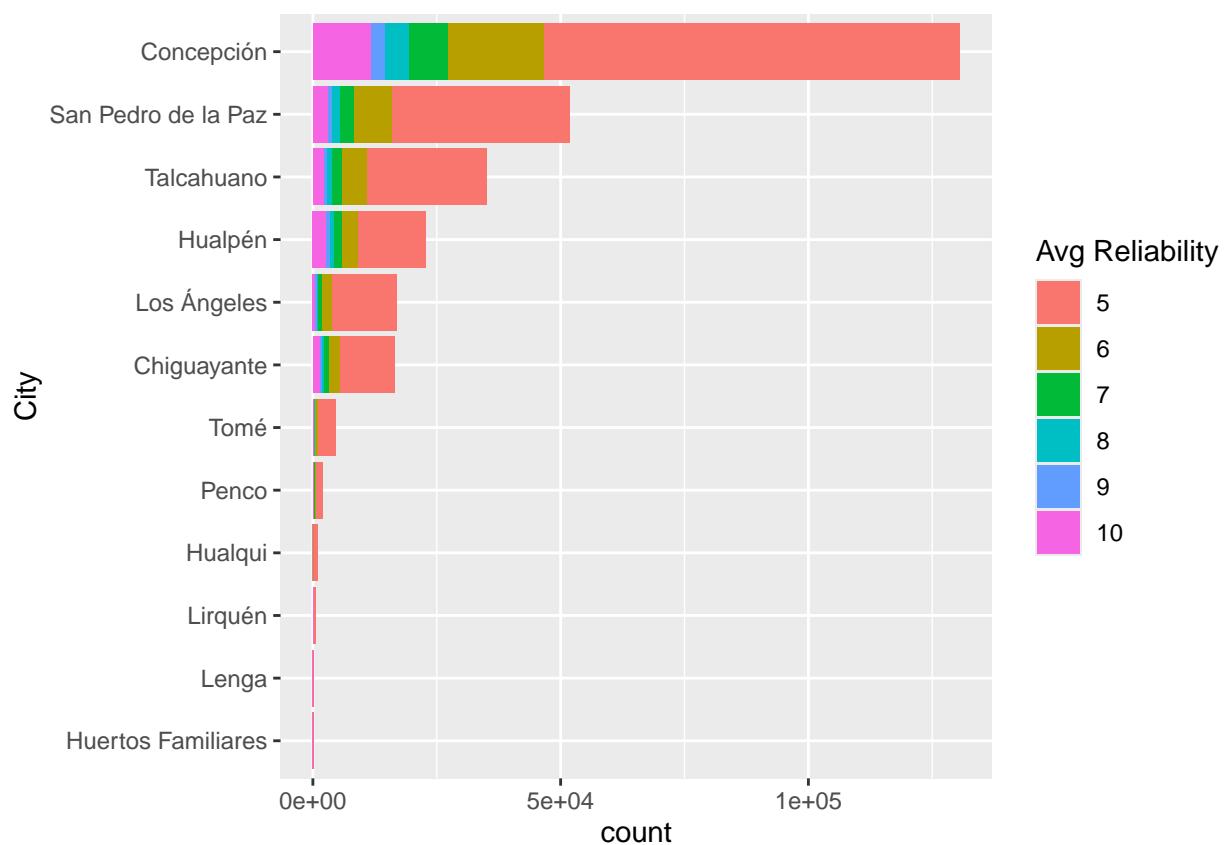
ROAD_CLOSED

Zoom: 15



Accidentes de tráfico

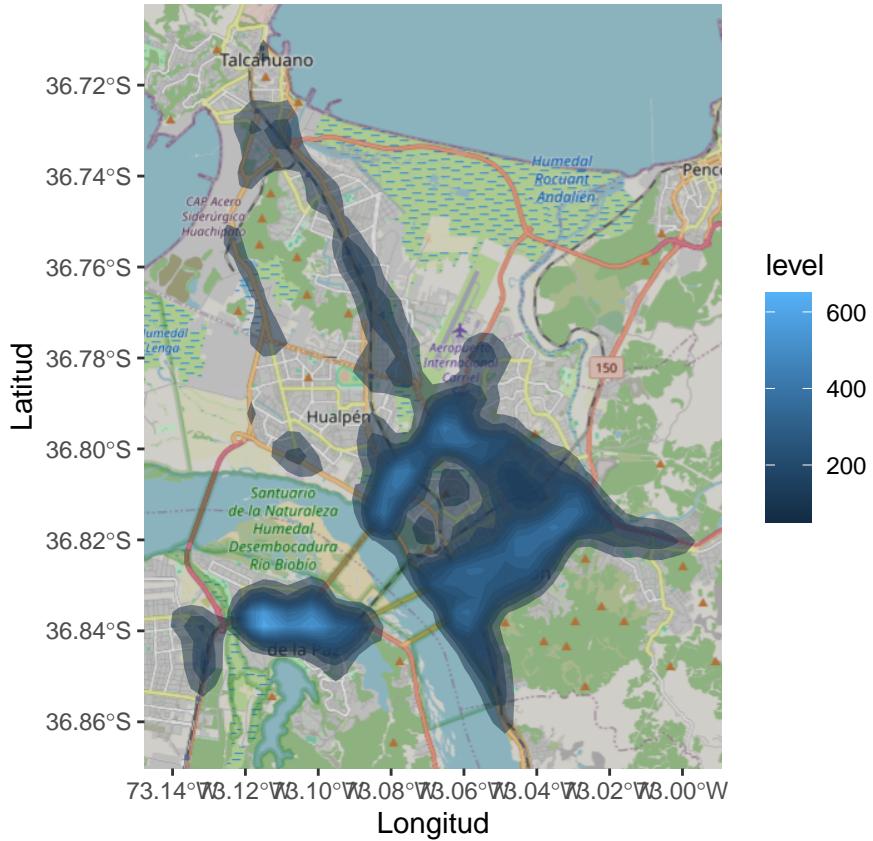
Distribución por ciudad



Distribución geográfica

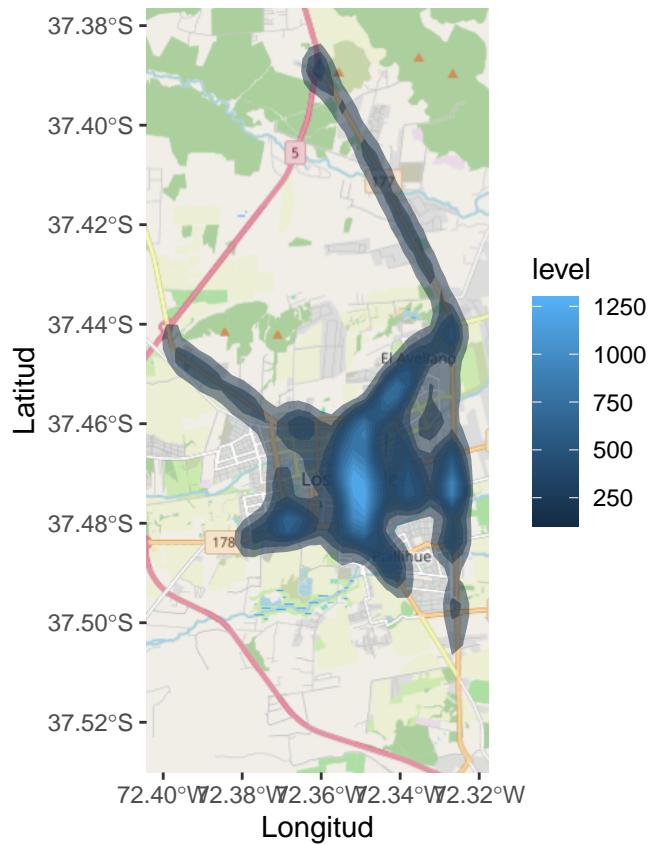
Concepción, San Pedro de la Paz, Talcahuano, Hualpén, Penco, Hualqui, Lirquén y Lenga

Zoom: 12



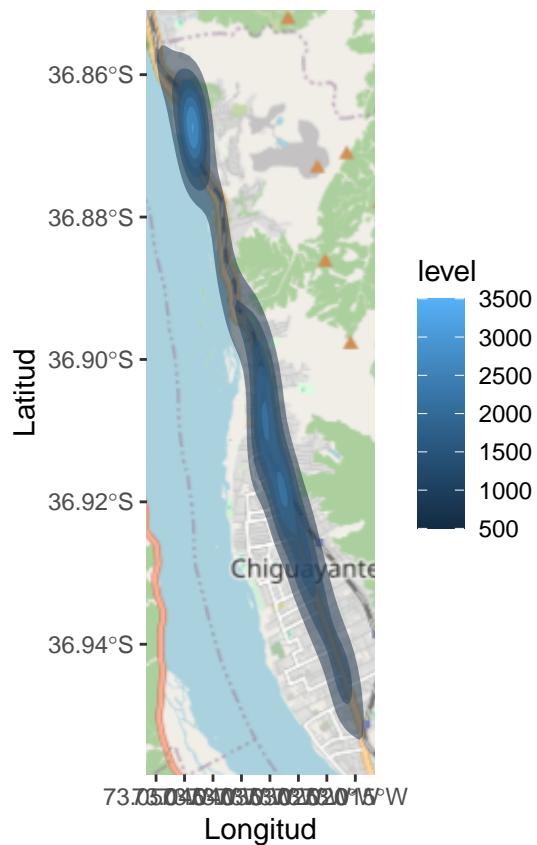
Los Ángeles

Zoom: 12



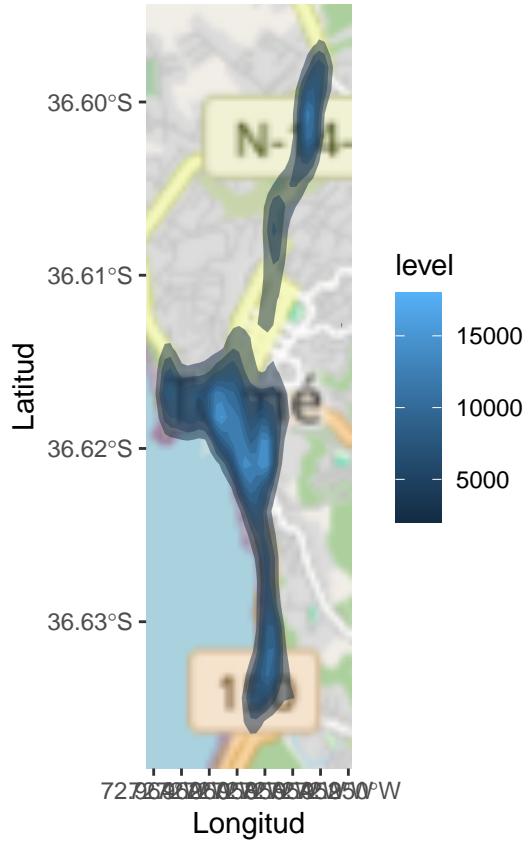
Chiguayante

Zoom: 12



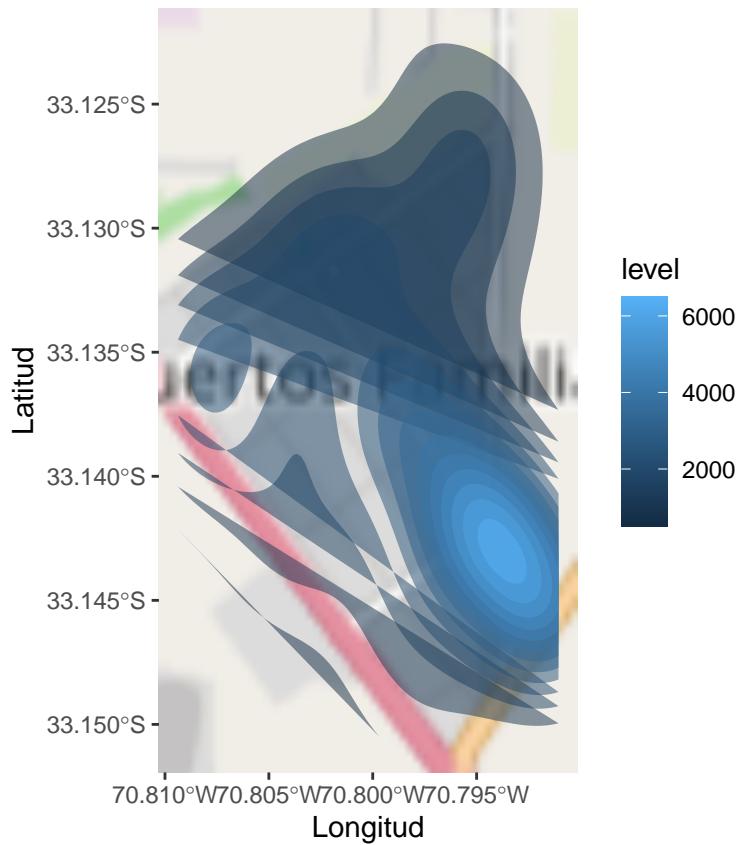
Tomé

Zoom: 12



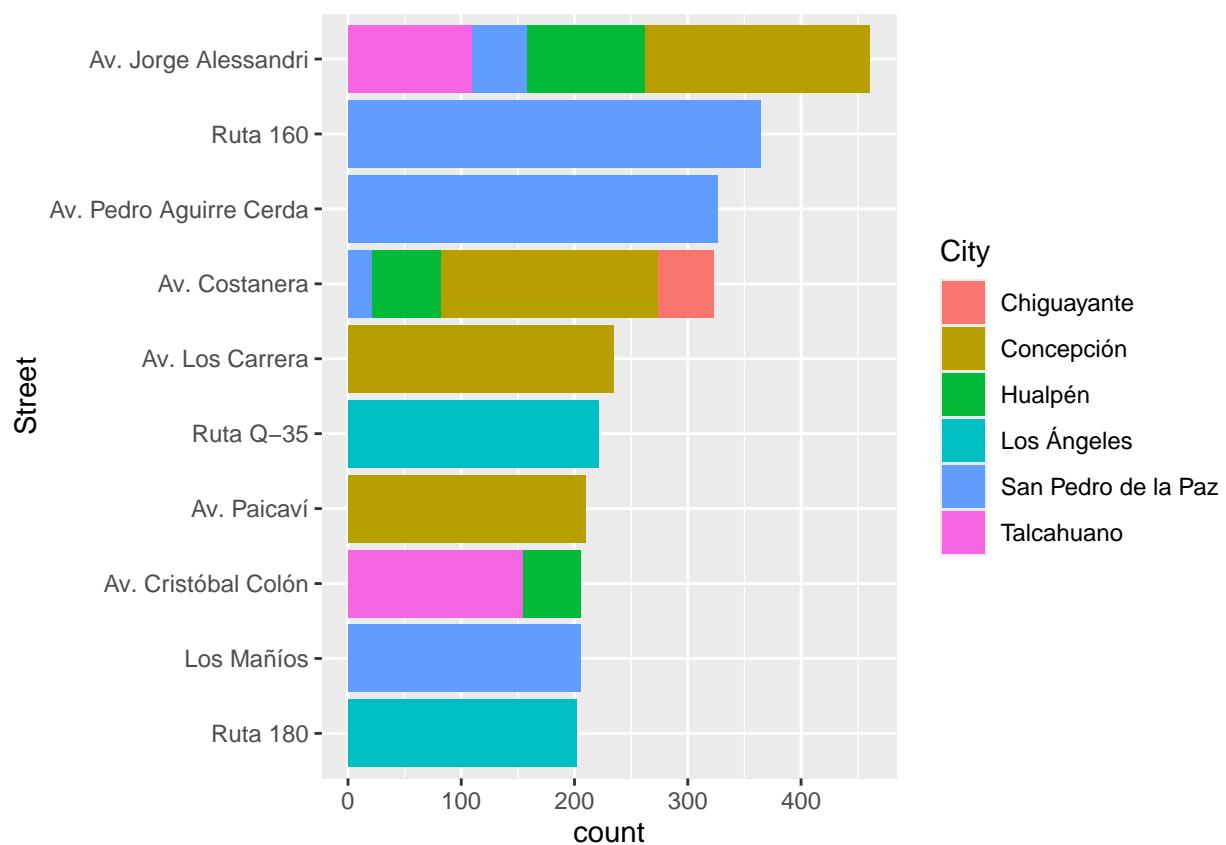
Huertos Familiares

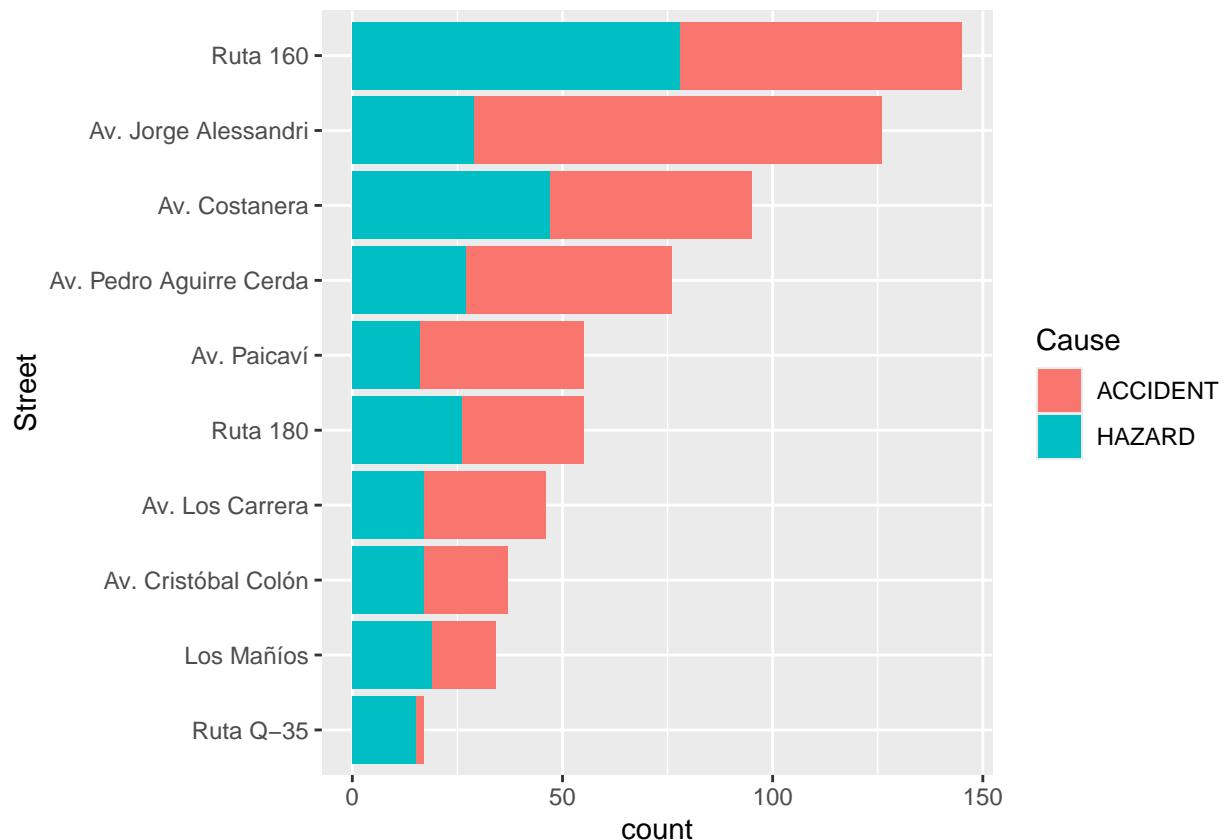
Zoom: 12



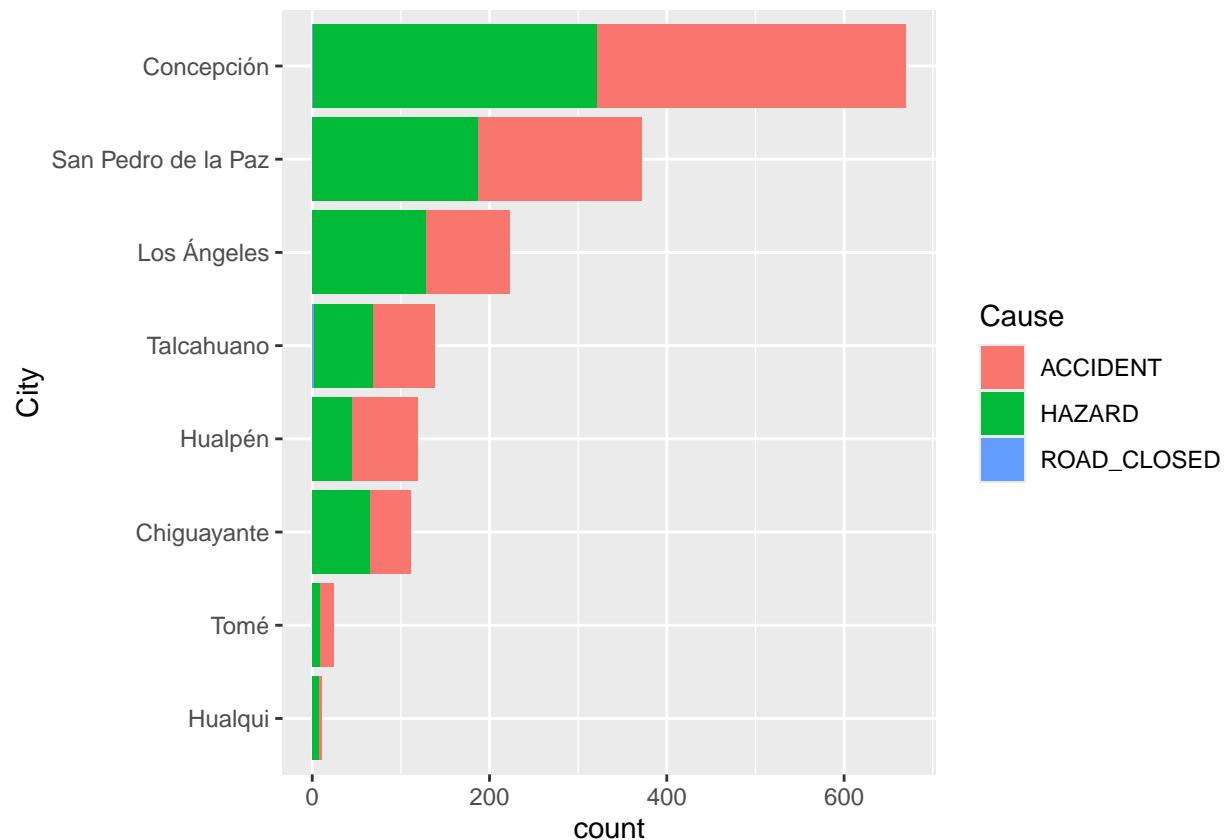
Alertas de Waze

Distribución por calles

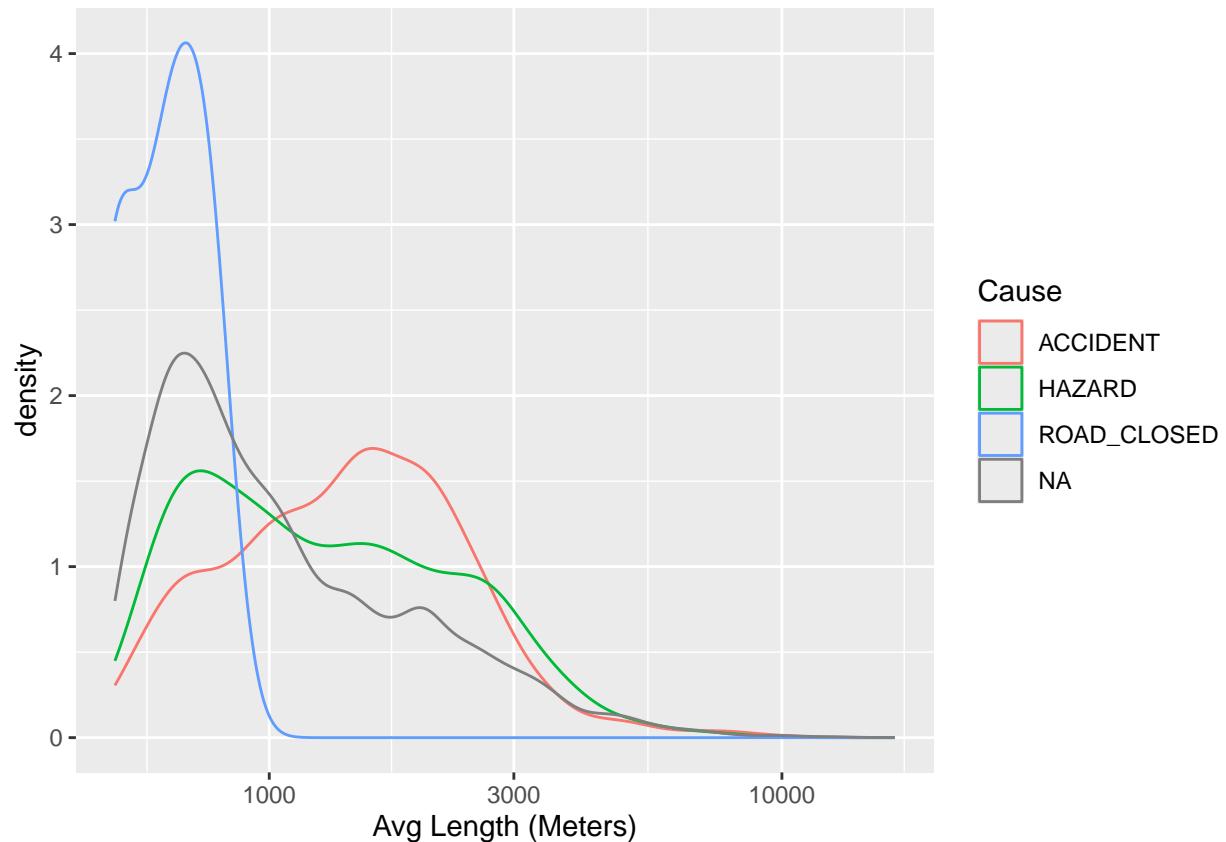




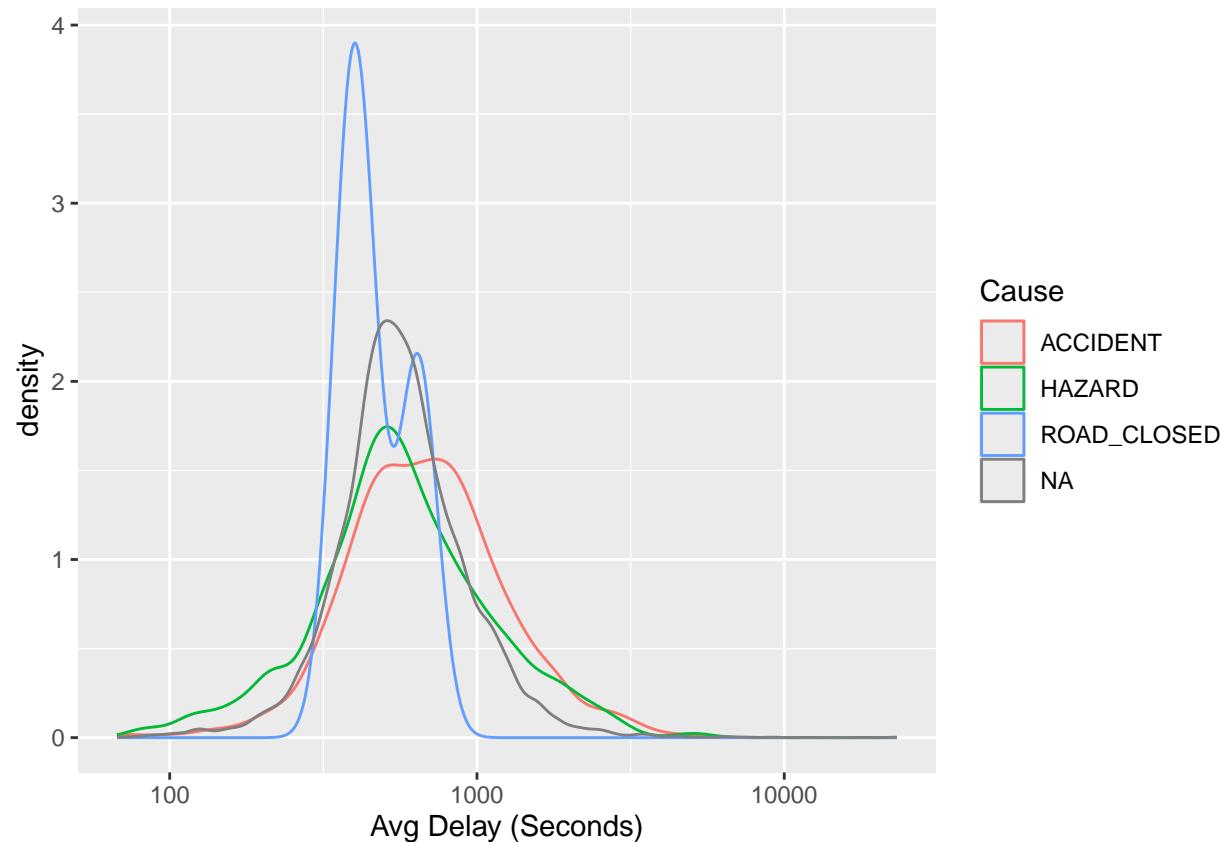
Distribución por ciudad



Distribución por largo promedio



Distribución por desfase promedio



Distribución por usuarios impactados

