

Tarea Final

Arriola, Miquelerena, Rovetta

2023-07-08

Introducción

Esto es un análisis descriptivo de los datos del tráfico de Montevideo, Uruguay.

Datos

Descripción general de los datos

Todos los datos fueron sacados de Catalogo de Datos Abiertos de gub.uy. En particular, los datos elegidos son los siguientes

- Conteo vehicular en las principales avenidas de Montevideo
- Velocidad promedio vehicular en las principales avenidas de Montevideo
- Ubicación de sensores de medición de conteo vehículos

Los tres dataset son mantenidos por la Intendencia de Montevideo.

Descripción de variables

Originalmente los datos vienen presentados de la siguiente forma:

Conteo vehicular en las principales avenidas de Montevideo

- **cod_detector**: Entero - ID de la cámara que monitorea un carril específico para detectar vehículos.
- **id_carril**: Entero - Número del carril monitoreado (1, 2, 3, ...).
- **fecha**: AAAA-MM-DD - Día en que se realizó la medición.
- **hora**: hh:mm:ss - Hora en que se realizó la medición.
- **dsc_avenida**: Texto - Nombre de la avenida donde se mide el tráfico.
- **dsc_int_anterior**: Texto - Nombre de la vía desde donde vienen los vehículos.
- **dsc_int_siguiente**: Texto - Nombre de la vía hacia donde se dirigen los vehículos.
- **latitud**: Float - Latitud del lugar de medición.
- **longitud**: Float - Longitud del lugar de medición.
- **volumen**: Entero - Cantidad de vehículos detectados en el carril en los últimos 5 minutos.
- **volumen_hora**: Entero - Cantidad de vehículos detectados en el carril en la última hora.

Velocidad promedio vehicular en las principales avenidas de Montevideo

- **cod_detector:** Entero - ID de la cámara que monitorea un carril específico para detectar vehículos.
- **id_carril:** Entero - Número del carril monitoreado (1, 2, 3, ...).
- **fecha:** AAAA-MM-DD - Día en que se realizó la medición.
- **hora:** hh:mm:ss - Hora en que se realizó la medición.
- **dsc_avenida:** Texto - Nombre de la avenida donde se mide el tráfico.
- **dsc_int_anterior:** Texto - Nombre de la vía desde donde vienen los vehículos.
- **dsc_int_siguiente:** Texto - Nombre de la vía hacia donde se dirigen los vehículos.
- **latitud:** Float - Latitud del lugar de medición.
- **longitud:** Float - Longitud del lugar de medición.
- **velocidad_promedio:** Entero - Promedio de las velocidades de los autos que circularon por el carril durante los últimos 5 minutos.

Ubicación de sensores de medición de conteo vehículos

- **dsc_avenida:** Texto - Nombre de la avenida donde se encuentra el sensor o cámara y donde se mide el tránsito.
- **dsc_int_anterior:** Texto - Nombre de la vía que forma el cruce desde donde vienen los vehículos.
- **dsc_int_siguiente:** Texto - Nombre de la vía que forma el cruce donde está el sensor. En general, el sensor se encuentra un poco antes de esta vía. El sentido de circulación será desde el cruce con **dsc_int_anterior** hacia el cruce con **dsc_int_siguiente**.
- **latitud:** Float - Coordenada que indica la latitud de la ubicación del sensor.
- **longitud:** Float - Coordenada que indica la longitud de la ubicación del sensor.

Sobre estos datos en particular, son *100 sensores* que se van cambiando de ubicación mes a mes

Base de Datos

Debido a que los datos utilizados están estrechamente relacionados y a su vez son sumamente masivos, hemos decidido utilizar una base de datos.

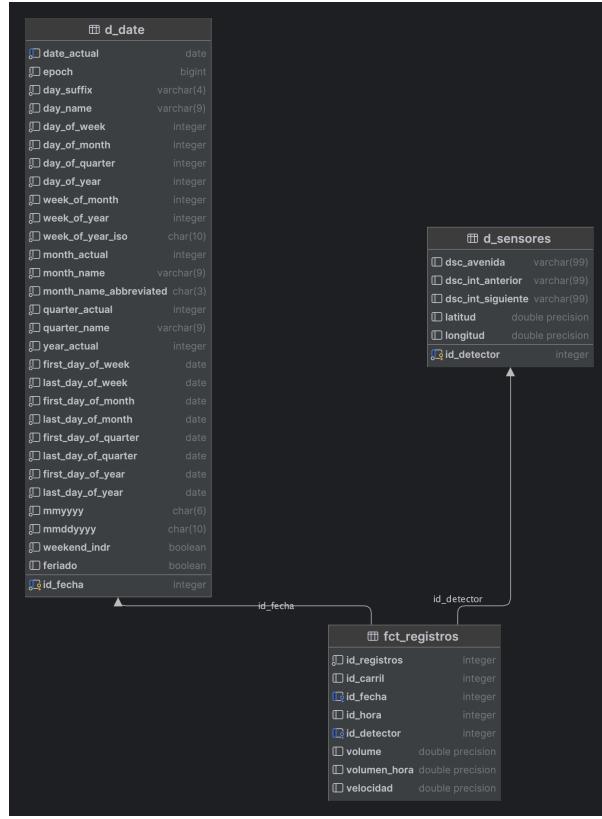


Figure 1: Diagrama de la base de datos

En la base de datos la tabla principal es **fct_registros**.

Así es como quedan los datos en las diferentes

Tabla: fct_registros

- Cantidad de datos: 85386695
- Variables de la tabla:
 - *id_registros* (Primary Key)
 - *id_carril*
 - *id_fecha* -> d_date(*id_fecha*) (Foreign Key)
 - *id_hora*
 - *id_detector* -> d_sensores(*id_detector*) (Foreign Key)
 - *volume*
 - *volumen_hora*
 - *velocidad*

Tabla: d_sensores

- Cantidad de datos: 273
- Variables de la tabla:
 - *id_detector* (Primary Key)
 - *dsc_avenida*

- dsc_int_anterior
- dsc_int_siguiente
- latitud
- longitud
- barrio

Tabla: d_date

- Cantidad de datos: 3652
- Variables de la tabla:
 - *id_fecha* (Primary Key)
 - date_actual
 - epoch
 - day_suffix
 - day_name
 - day_of_week
 - day_of_month
 - day_of_quarter
 - day_of_year
 - week_of_month
 - week_of_year
 - week_of_year_iso
 - month_actual
 - month_name
 - month_name_abbreviated
 - quarter_actual
 - quarter_name
 - year_actual
 - first_day_of_week
 - last_day_of_week
 - first_day_of_month
 - last_day_of_month
 - first_day_of_quarter
 - last_day_of_quarter
 - first_day_of_year
 - last_day_of_year
 - mmYYYY
 - mmddYYYY
 - weekend_indr
 - feriado

Análisis exploratorio

Sobre la ubicacion de los datos

Para empezar mostraremos la superficie de Montevideo alcanzada por la informacion de los sensores de la intendencia. Esta informacion no alcanza a todos los barrios , sino que solamente alcanza los siguientes barrios

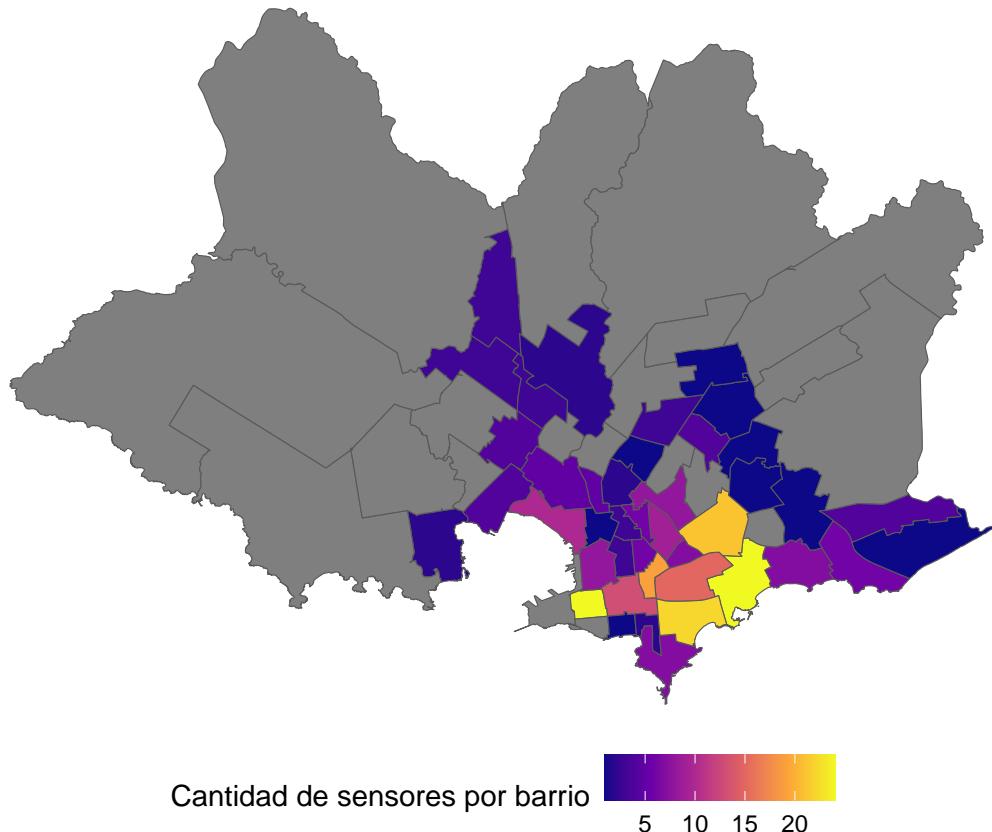
```
## # A tibble: 42 x 2
##   barrio          cant_de_sensores
```

```

##      <chr>          <int>
## 1 BUCEO              24
## 2 CENTRO             24
## 3 POCITOS            22
## 4 UNION               21
## 5 TRES CRUCES        19
## 6 PQUE. BATLLE - V. DOLORES 15
## 7 CORDON              13
## 8 CAPURRO - BELLA VISTA 10
## 9 LARRAÑAGA            9
## 10 AGUADA              8
## # i 32 more rows

```

En el siguiente mapa se puede visualizar



““

Sobre el tiempo abarcado de los datos

Los datos fueron recolectados cada 5 minutos, por diferentes sensores de la intendencia de montevideo. Comienzan desde Enero del 2021 hasta Mayo del 2023.

Velocidad

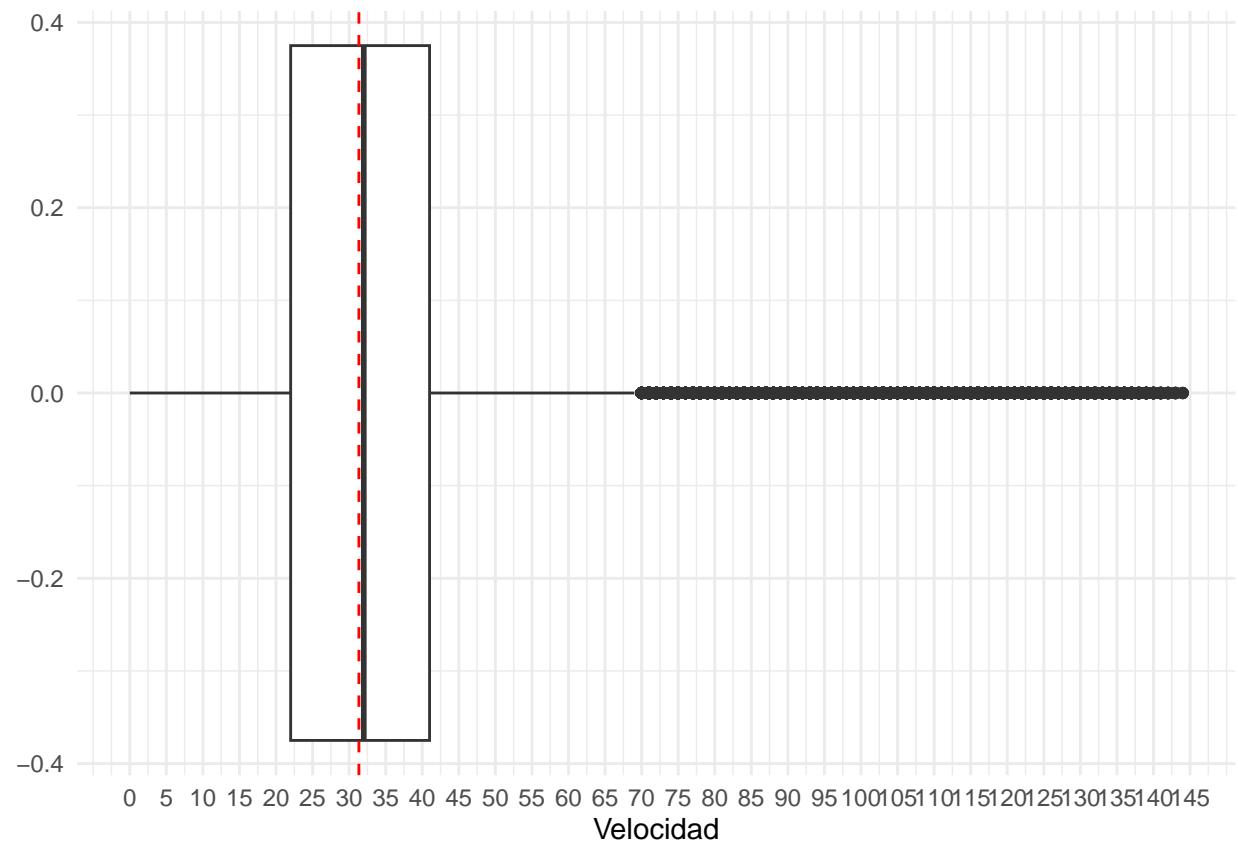
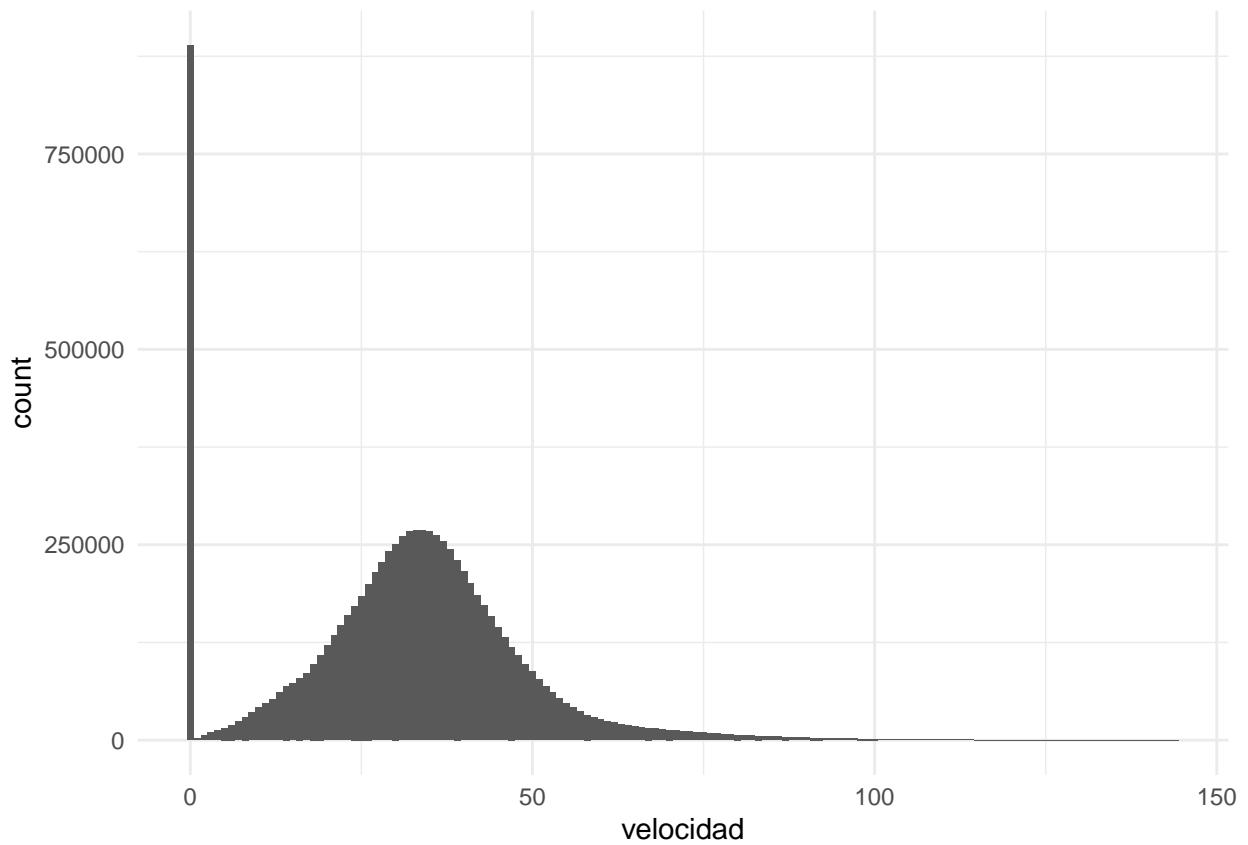


Figure 2: Boxplot de Velocidad



Volumen

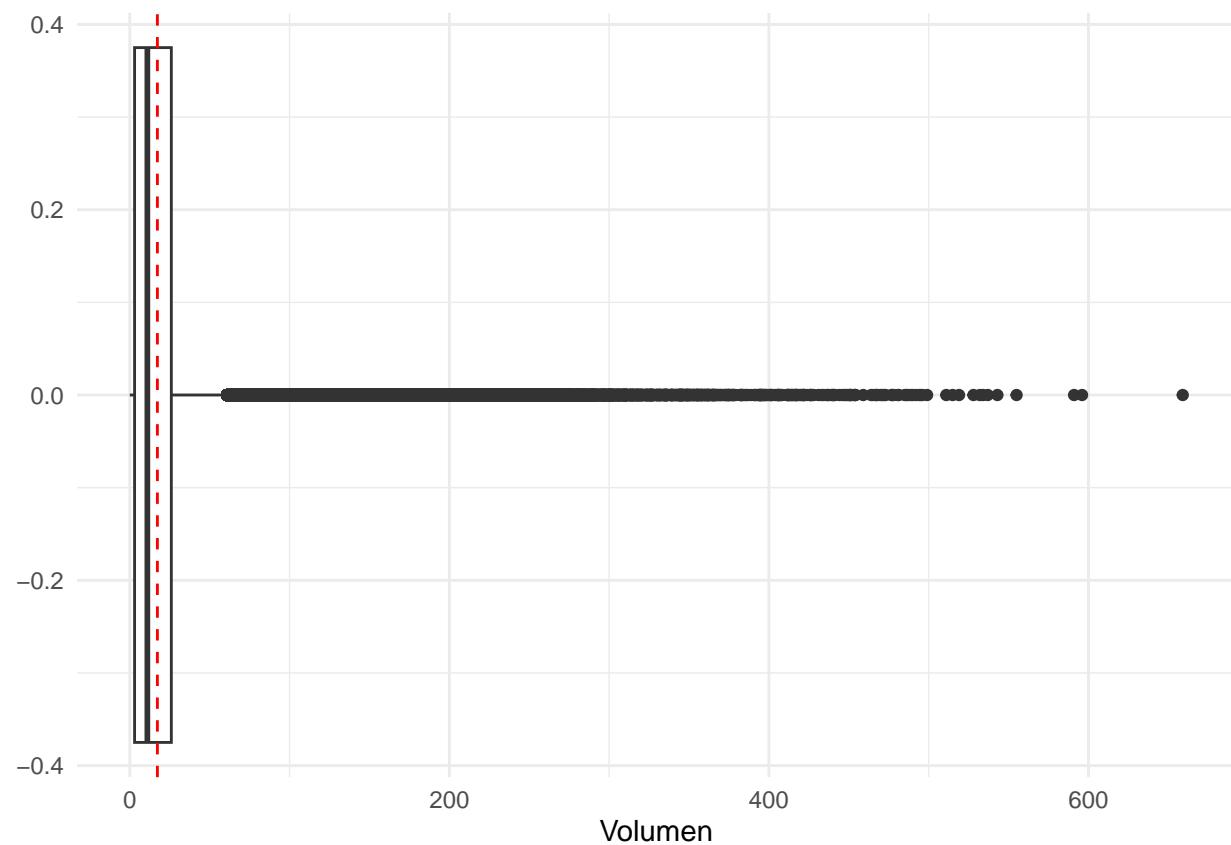
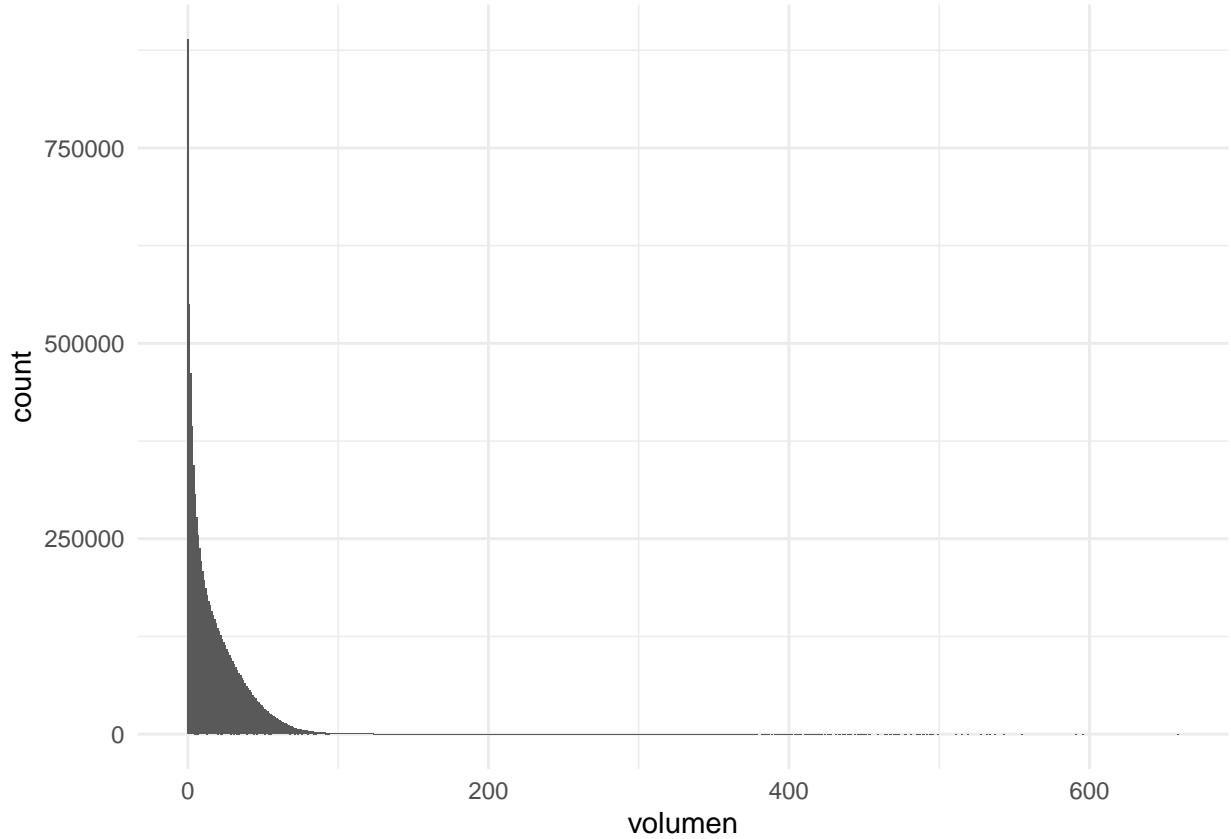


Figure 3: Boxplot de Volumen

Se puede ver que volumen es muy dispar en su distribucion



Muestra aleatoria de volumen y velocidad, agregados por dia de la semana

Vamos a analizar la relación entre el volumen de tráfico, la velocidad, el día de la semana y el barrio donde se encuentra el sensor

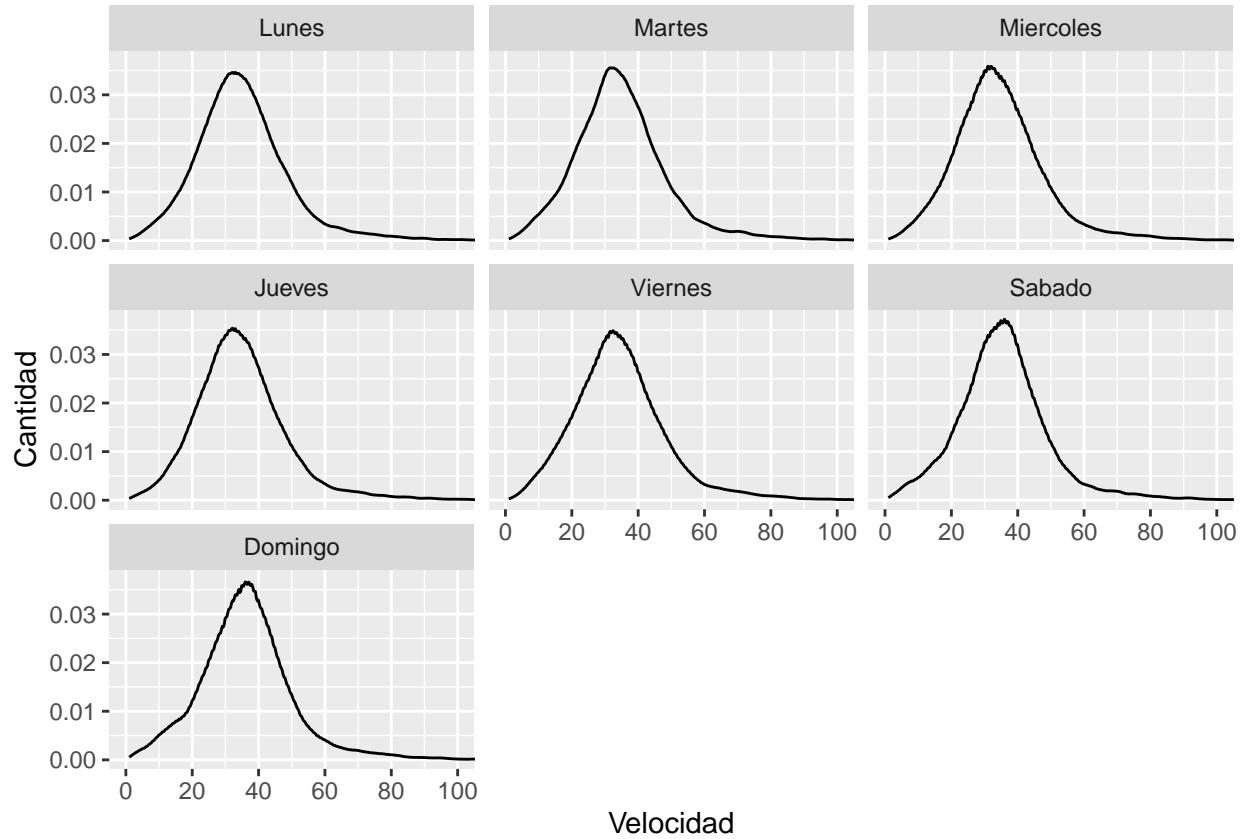


Figure 4: Densidad de Velocidad por Dia de la semana

Se puede ver que la distribucion de velocidad casi no cambia por dia de la semana

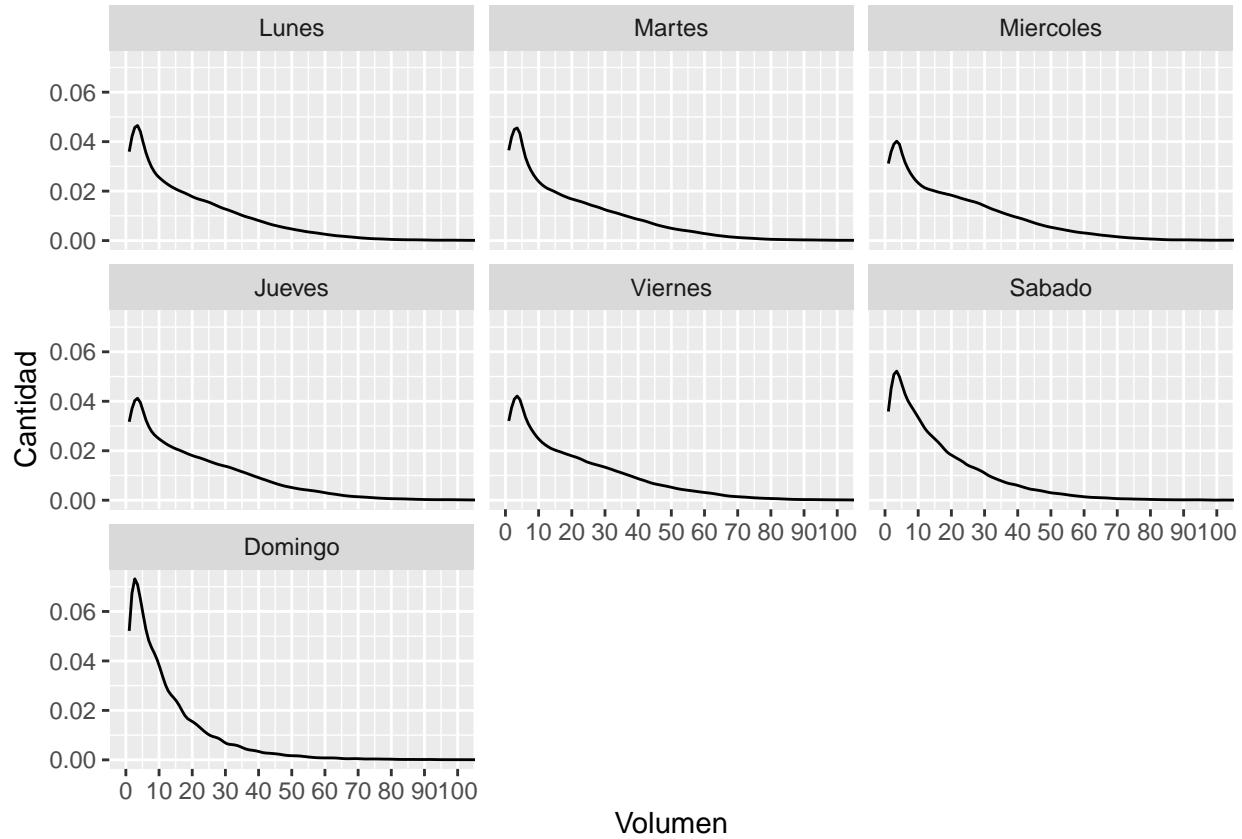


Figure 5: Densidad de Volumen por Dia de la semana

Se puede observar como las concentraciones mas pequeñas son mas frecuentes los fines de semana, en especial los domingos

Preguntas de investigación

1. ¿Existe alguna correlación entre el volumen y la velocidad?
2. ¿Cuáles son las calles con los mayores promedios de velocidad en Montevideo? ¿Con qué frecuencia se cometen excesos de velocidad?
3. ¿Cómo va variando el volumen y velocidad medidos a través del TIEMPO?

¿Existe alguna correlación entre el volumen y la velocidad?

Haremos un grafico de puntos para visualizarlo. Para los datos usaremos una muestra aleatoria de toda la base de datos

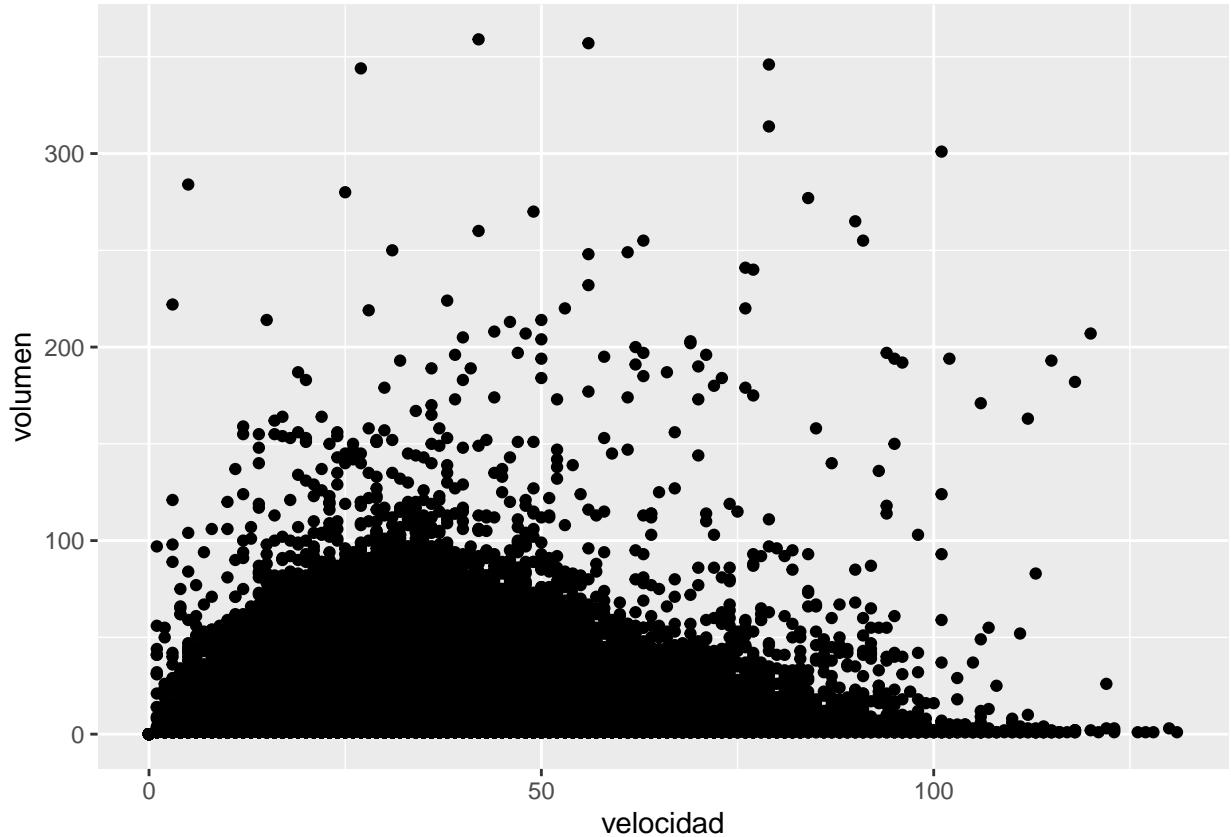


Figure 6: Grafico de puntos de velocidad y volumen

Definitivamente no hay una relacion lineal entre velocidad y volumen

```
##           velocidad     volumen
## velocidad 1.0000000 0.1671053
## volumen    0.1671053 1.0000000
```

¿Cuáles son las calles con los mayores promedios de velocidad en Montevideo?
¿Con qué frecuencia se cometen excesos de velocidad?

Pasemos a investigar las calles con mayor promedio de velocidad.

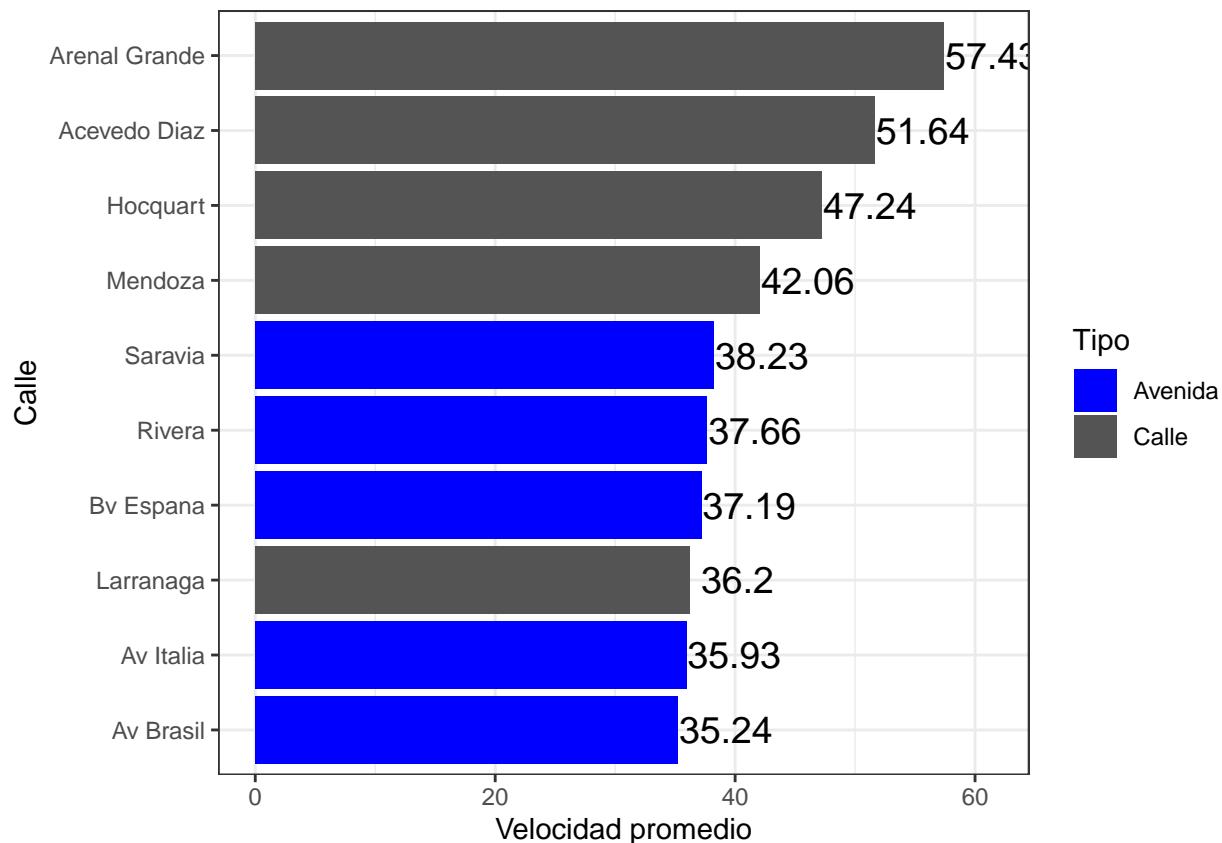


Figure 7: Calles con mayor promedio de velocidad

El siguiente gráfico nos muestra que las 3 calles con mas velocidad en promedio superan el máximo de 45km/h siendo este la velocidad máxima de circulación reglamentaria. La avenida más rápida en promedio no alcanza el máximo de velocidad permitido

En promedio de velocidad circulacion los conductores son prudentes, aun asi vemos con que frecuencia se comenten excesos de velocidad.

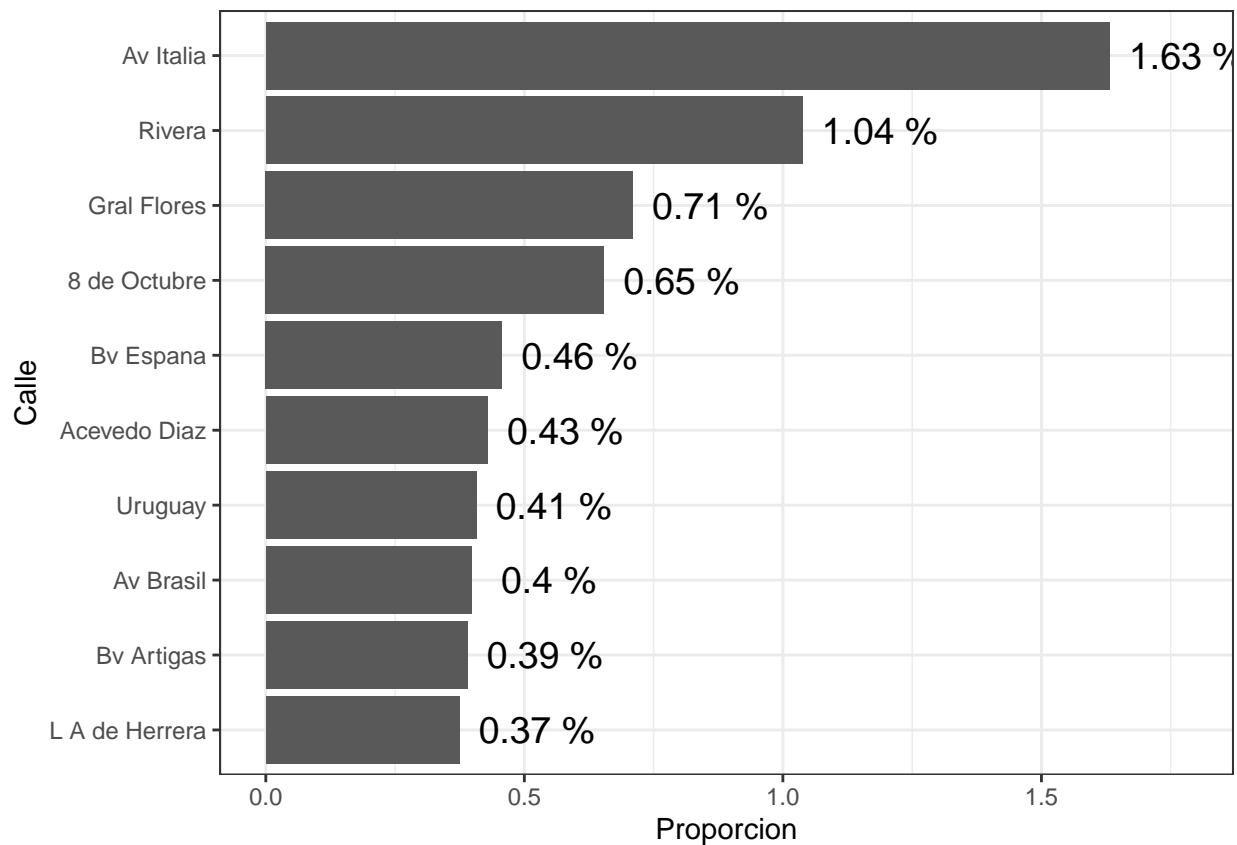
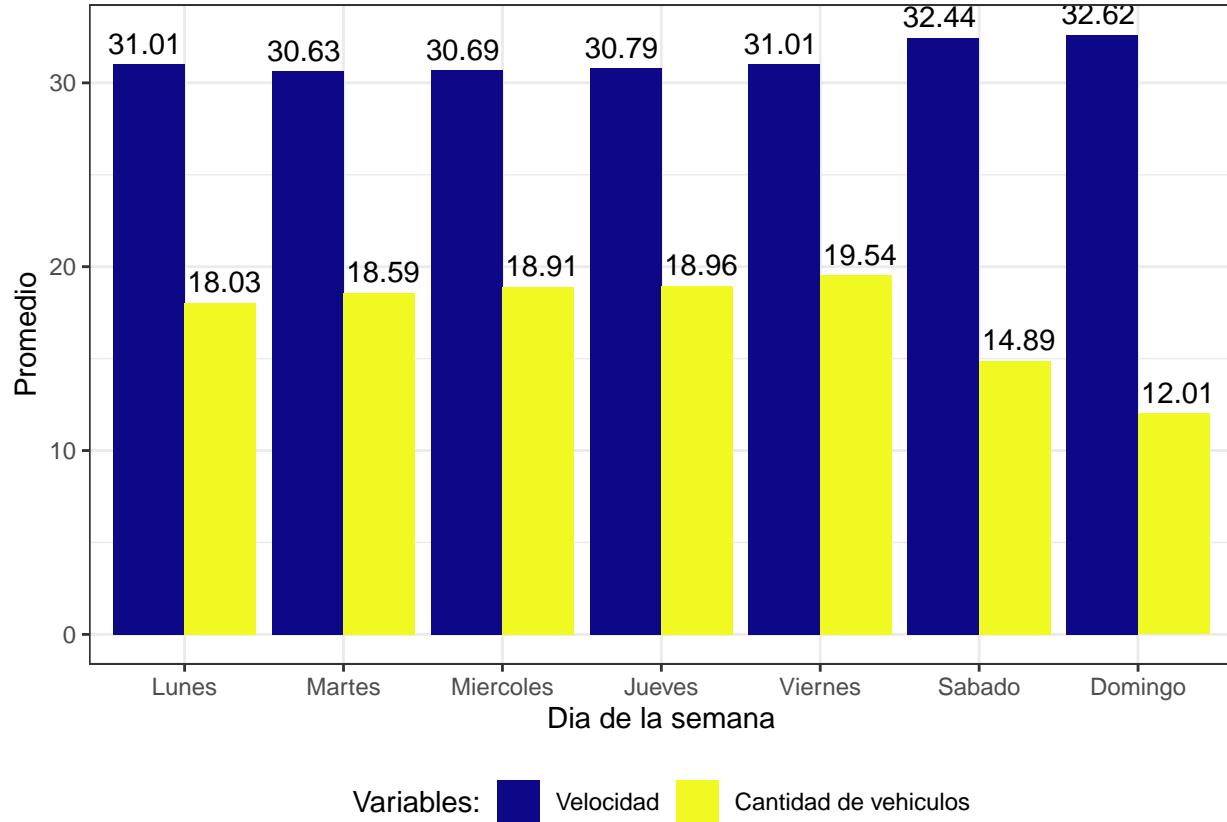


Figure 8: Es netamente nula la proporción de excesos de velocidad, podemos decir el los conductores en general respetan los límites de velocidad

¿Cómo va variando el volumen y velocidad medidos a través de el tiempo?

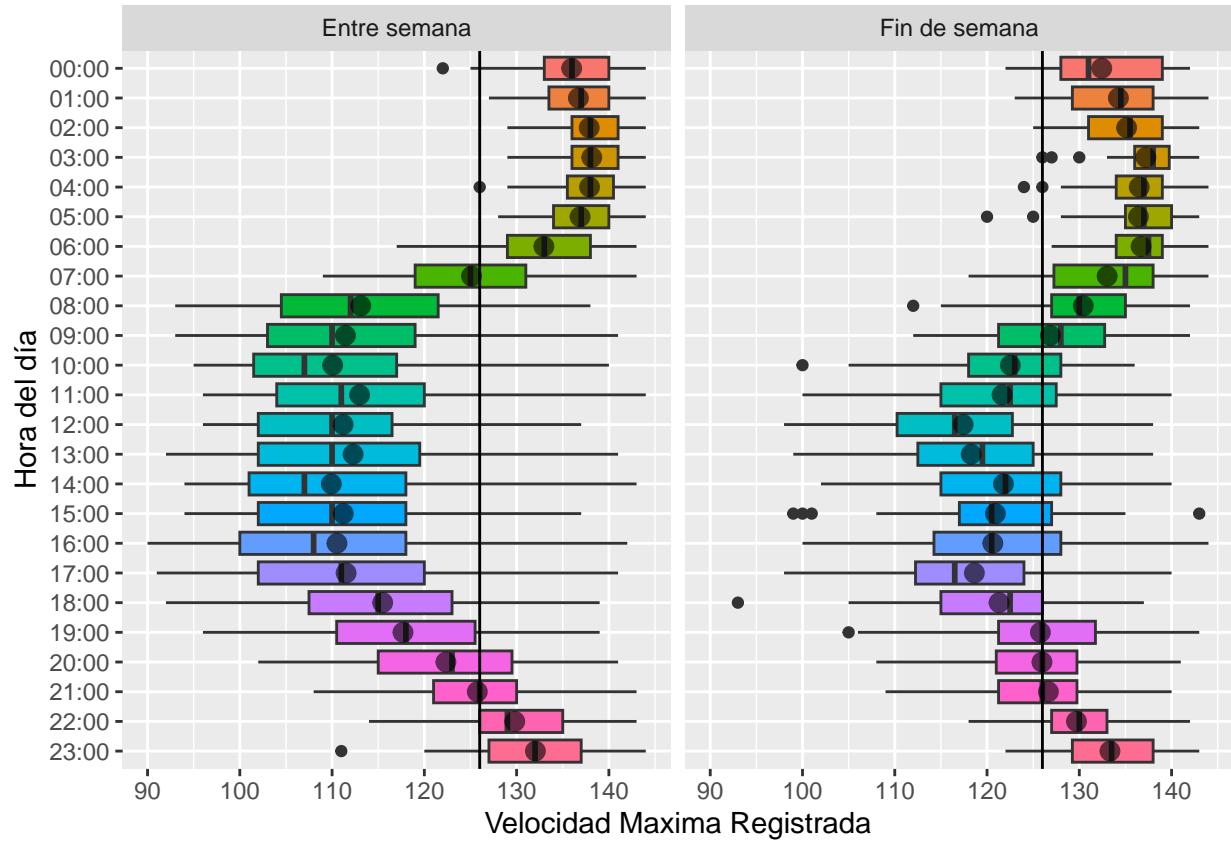
Observemos el volumen y la velocidad promedio de todos los días de la semana



En el gráfico se puede observar:

- La velocidad media se mantiene casi constante durante toda la semana, salvo una leve variación positiva los fines de semana.
- El volumen medio de vehículos detectados por los radares va variando en la semana, alcanzando su pico los viernes. También es considerablemente inferior los fines de semana.

Si bien previamente observamos leves excesos de velocidad, dentro de esos excesos se registran velocidades máximas de más de 100km/h, veamos esos máximos pro franja horaria según si fue registrado en un fin de semana o no



Entre las 00:00 y las 08:00 es donde se alcanzan picos de casi 140 km/h, luego de ese rango se reduce la tendencia por ser horario laboral para luego crecer a medida de que cae la noche. Durante el fin de semana las velocidades máximas durante el horario laboral de lunes a viernes es totalmente diferente, los conductores durante el fin de semana aprovechan la baja de volumen para circular mas rápido.

Si entramos en cada año vemos como crece la densidad de velocidad máxima registrada de lunes a viernes,

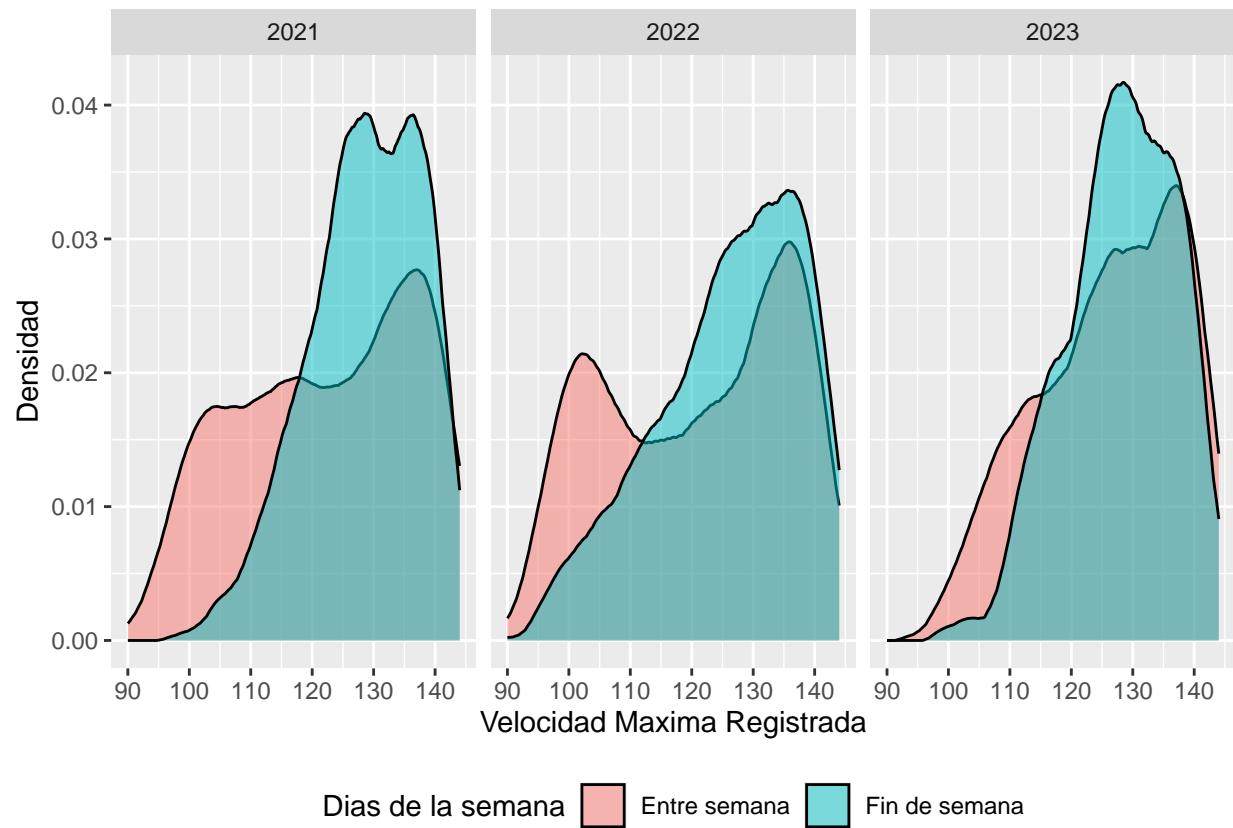


Figure 9: 2023 no esta completo, vamos a profundizar mas en nuestro modelo predictivo

Ahora la densidad de velocidad durante los dias siempre se encuentran en un rango de $(0,60]$

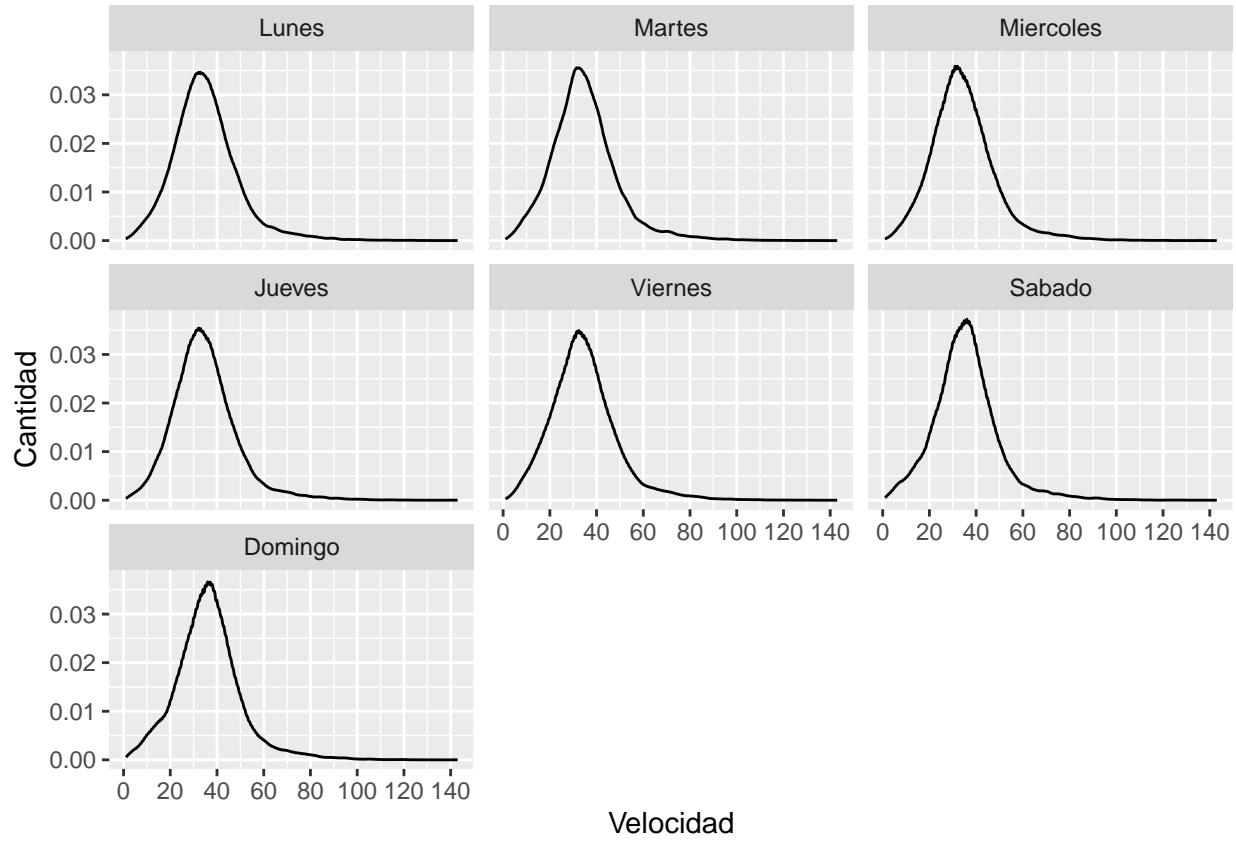


Figure 10: Densidad de Velocidad

Con el volumen de trafico, hay una gran diferencia el domingo comparado con el resto de los dias, en todos los dias la mayoria de volumen se concentra en el rango $(0,10]$ donde en dicho rango alcanza su pico de circulacion en todo Montevideo.

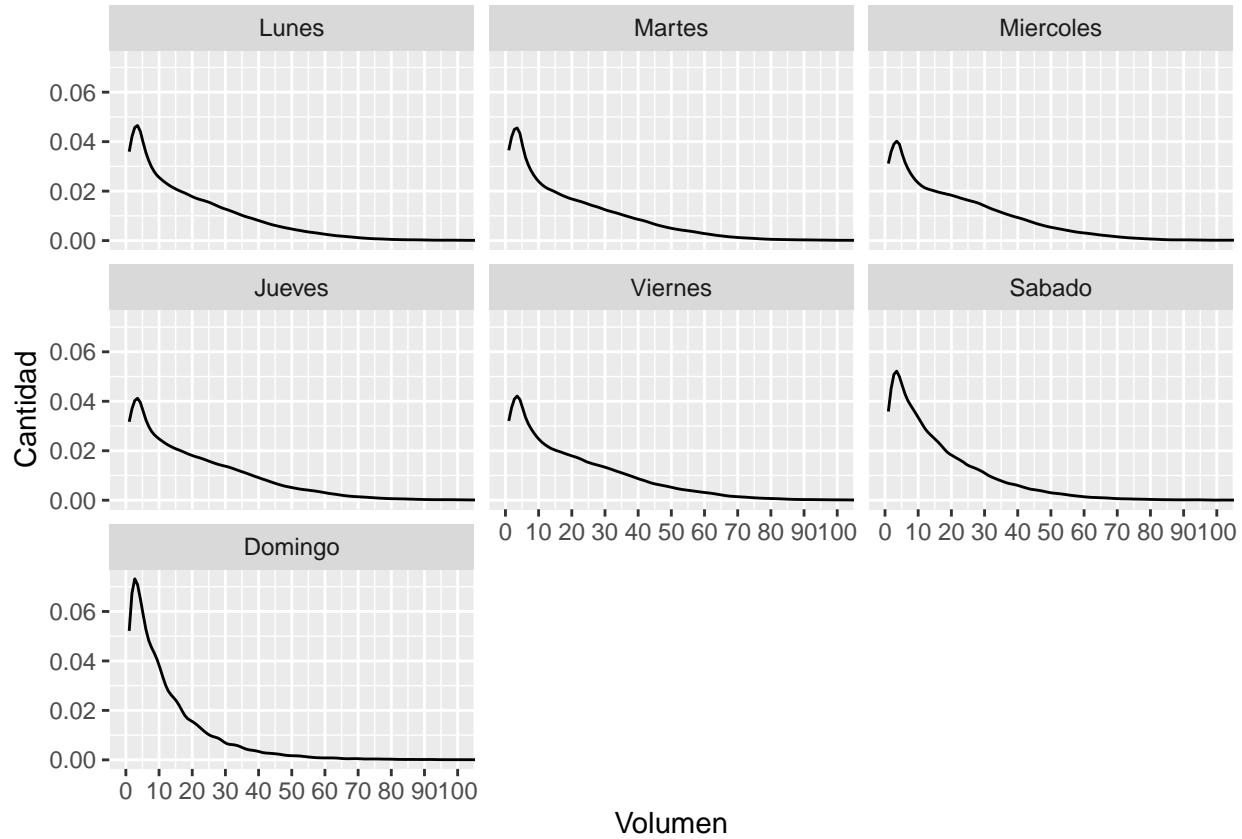
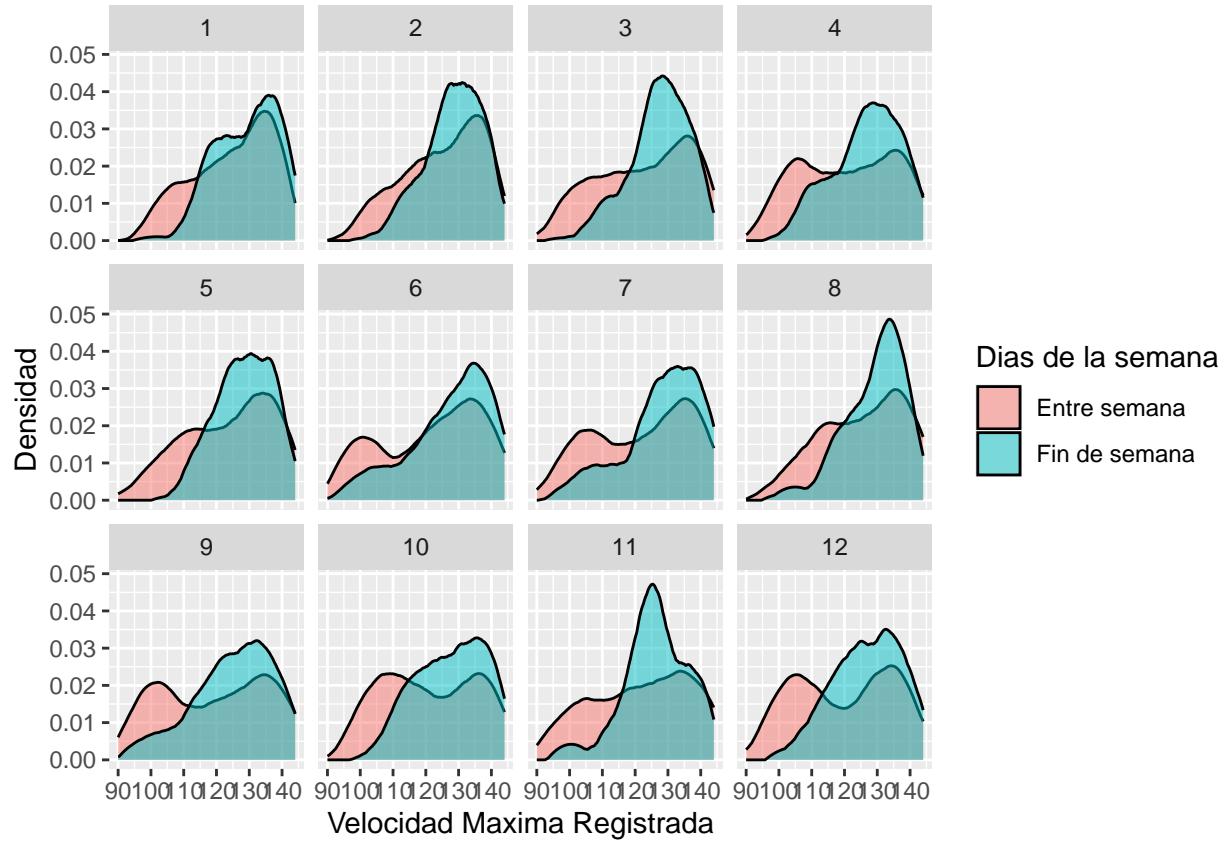


Figure 11: Densidad de Volumen



Se observa un pico en Agosto Noche de la nostalgia?

```
##      count month_actual year_actual
## 1 3830149          1       2022
## 2 3898585          1       2023
## 3 3714726          2       2021
## 4 3374893          2       2022
## 5 3455550          2       2023
## 6 3908628          3       2021
## 7 3746326          3       2022
## 8 3879877          3       2023
## 9 3759551          4       2021
## 10 3620380         4       2022
## 11 3930695         5       2021
## 12 2701766         5       2023
## 13 3873954         6       2021
## 14 3731149         6       2022
## 15 4078328         7       2021
## 16 3679998         7       2022
## 17 4032694         8       2021
## 18 3857623         9       2021
## 19 2501798         9       2022
## 20 3748052        10      2021
## 21 4072685        10      2022
## 22 3979504        11      2021
## 23 4009784        12      2021
```

Haciendo una observacion de los meses cubiertos por los datos, nos hemos dado cuenta que Agosto solo esta representado por los datos del 2021.

```
## Coordinate system already present. Adding new coordinate system, which will
## replace the existing one.
```

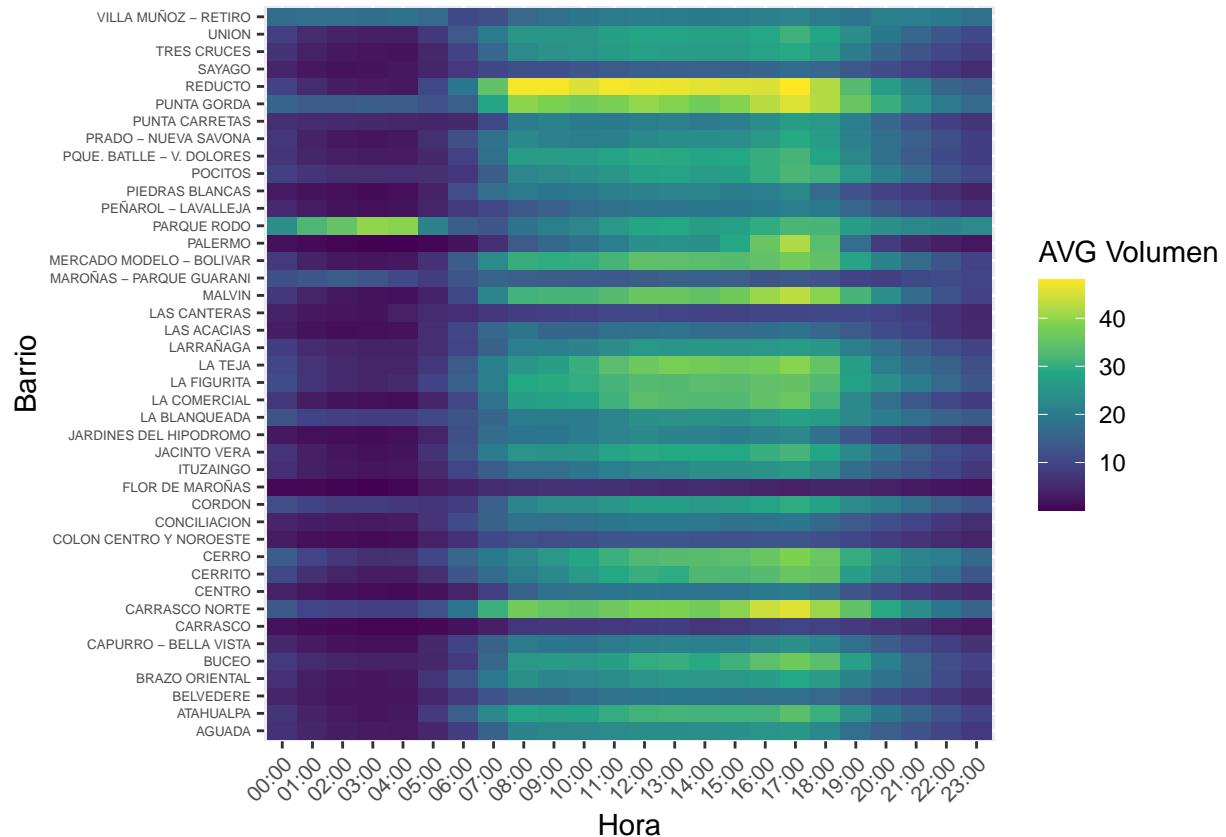
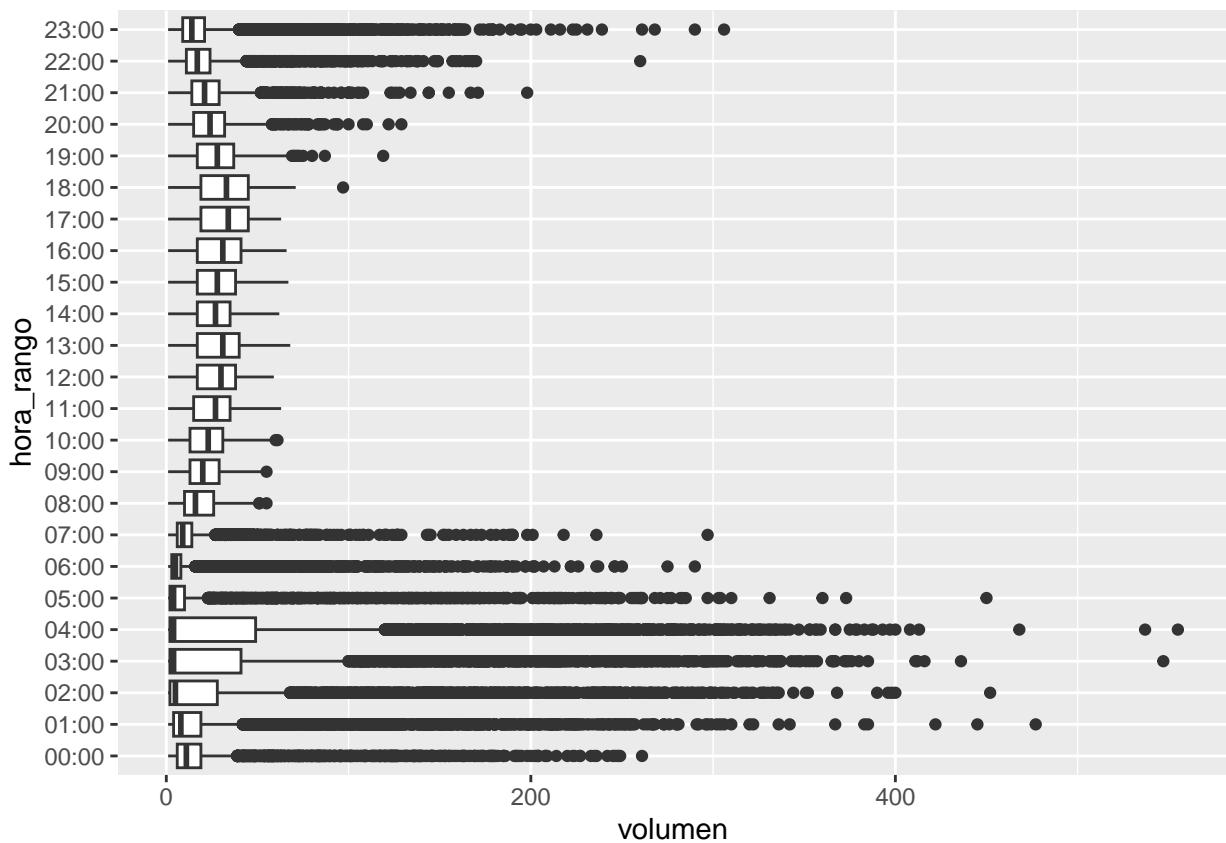


Figure 12: Mapa de calor de volumen maxima por barrio por rango de hora

En Parque Rodo hay valores altos durante la noche, que lo diferencia del resto de los barrios.



Resultados interesantes

[Presentar los resultados más relevantes obtenidos durante el análisis exploratorio.]

Modelo estadístico

Predicir velocidad promedio de un sensor

Que velocidad promedio se espera según - es fin de semana - es día o noche - volumen promedio

El modelo se aleja alrededor de esa cantidad de kilometros de los datos reales

Random Forest

El modelo le erra por esa cantidad de kilometros/hora en promedio

Este grafico muestra que variables son mas significativas para la predicción del promedio de velocidad

Predicciones

[Presentar las predicciones realizadas por el modelo.]

Interpretación de resultados

[Interpretar los resultados obtenidos del modelo estadístico.]

Aplicación Shiny

Descripción

La aplicación consta de 4 partes. Un mapa con la ubicación de los semáforos, un análisis univariado, otro multivariado y el modelo. La aplicación resume todo lo hecho anteriormente. Enlace

Comentarios finales

Hallazgos principales

[Resumir los principales hallazgos del proyecto.]

Posibles extensiones

[Discutir posibles extensiones o mejoras para el proyecto.]

Referencias

WICKHAM, HADLEY. 2023. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* O'REILLY MEDIA.