

Ausentismo_RezagosEstacionalidad

Ricardo Ortiz, Jeisson Rodriguez & Jaime Santa

2025-04-26

En el presente análisis se examina el comportamiento del ausentismo laboral a partir del conjunto de datos “Absenteeism at Work” disponible en el repositorio de la UCI Machine Learning. Para identificar patrones temporales y tendencias generales, se implementan técnicas de suavizamiento como la aproximación mediante promedios móviles. Además, se exploran los rezagos (lags) para evaluar la influencia de valores pasados en la dinámica actual del ausentismo, buscando detectar posibles relaciones de dependencia temporal. Finalmente, se realiza un análisis de estacionalidad para determinar si existen variaciones sistemáticas asociadas a ciclos recurrentes, como días de la semana, meses o estaciones del año, que afecten la frecuencia del ausentismo. Estas aproximaciones permiten construir una base sólida para entender mejor el fenómeno y orientar futuros modelos predictivos o estrategias de intervención en el contexto laboral.

A continuación se procede a cargar los datos de ausentismo laboral desde un archivo local, descargado directamente de: <https://archive.ics.uci.edu/dataset/445/absenteeism+at+work>. Se imprime el encabezado para ver el contenido del dataset.

Al revisar los datos se identifica que la variable de interés que puede ser apropiada para hacer el análisis es el tiempo de ausentismo en horas, disponible en la variable *Absenteeism.time.in.hours*. La cuál es la variable objetivo del dataset según la descripción del mismo.

Análisis Descriptivo

En primer lugar vemos algunas estadísticas descriptivas básicas para entender el comportamiento de la variable seleccionada:

Desviación estándar

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.000   2.000   3.000   6.924   8.000  120.000
## [1] 13.331
```

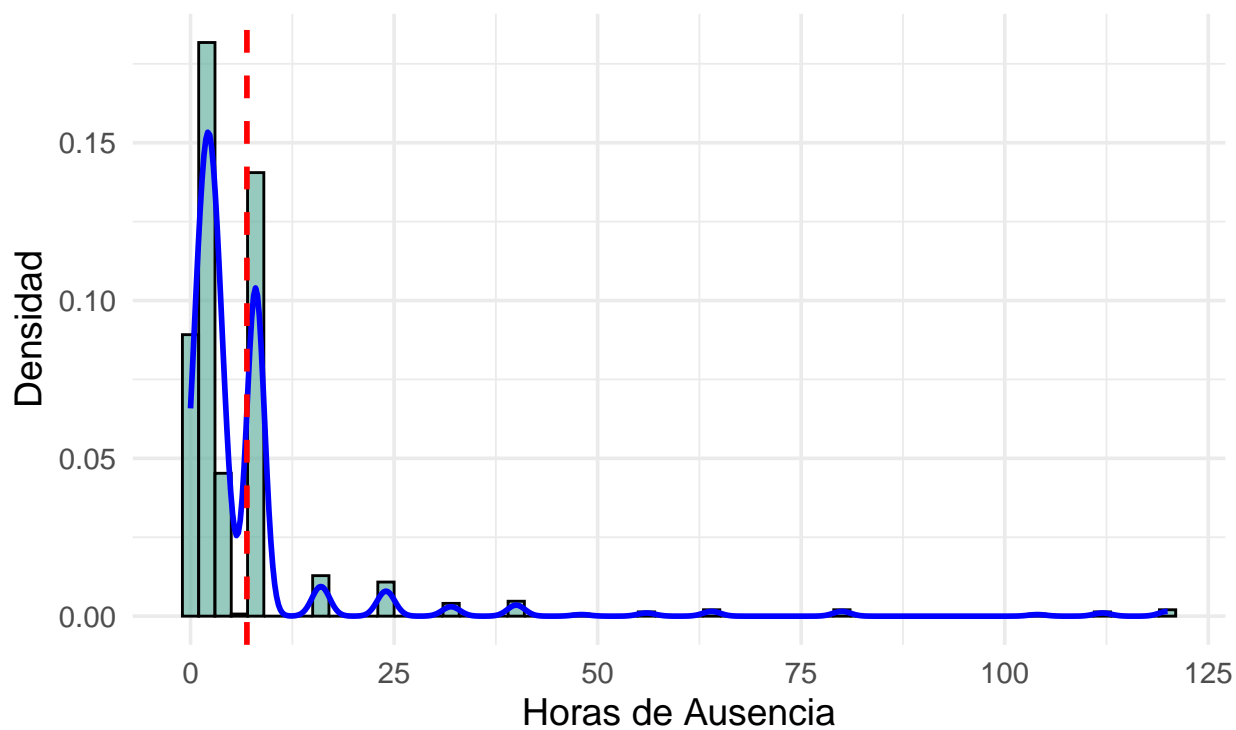
Acá vemos que el tiempo máximo de ausentismo es de 120 horas, con una mediana de 3 horas y un promedio de 7 horas aproximadamente. La desviación estándar arroja 13.3, esto indica que hay una variabilidad considerable en los datos, los datos están relativamente dispersos alrededor de la media.

Ahora vemos como se distribuyen los las hora de ausencia:

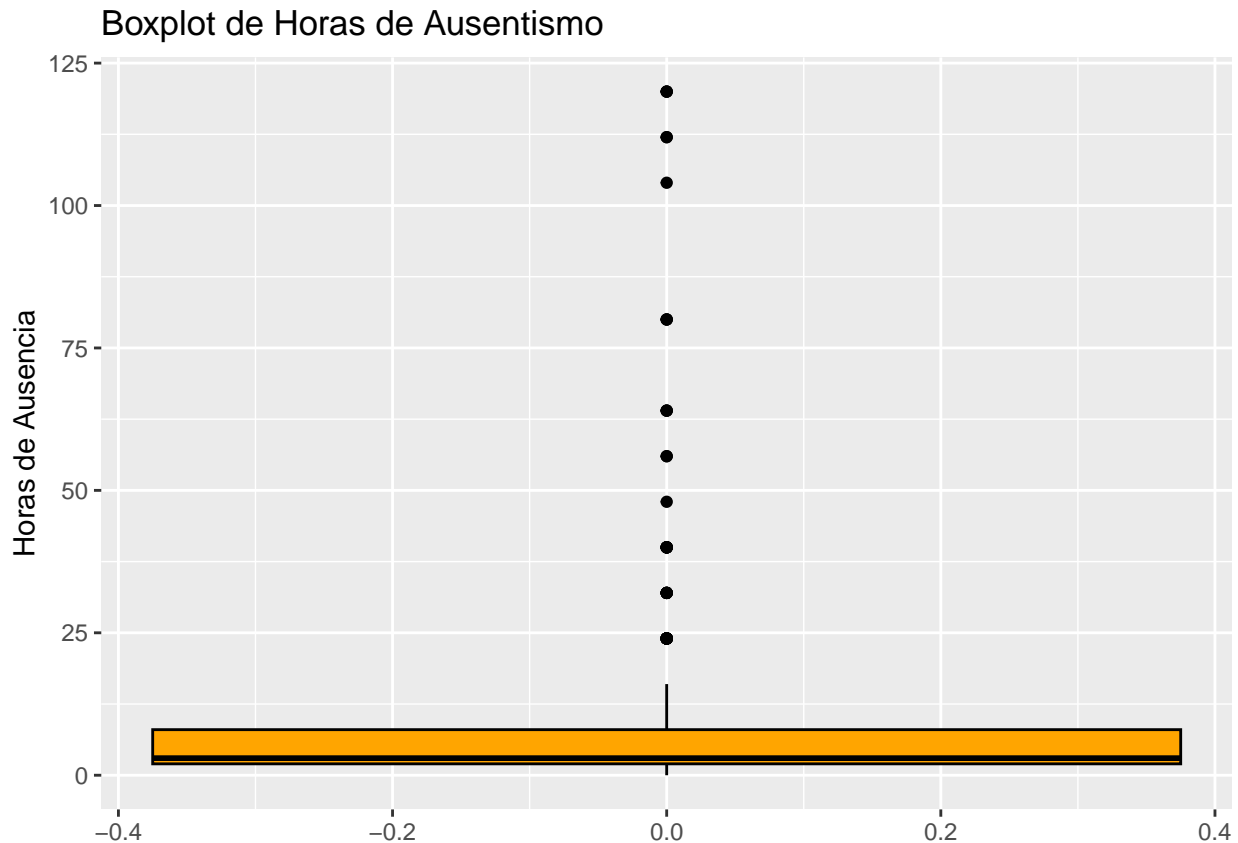
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

Distribución de Horas de Ausentismo Laboral

Media = 6.92 horas



Del gráfico se puede observar que hay una asimetría positiva y que algunos valores de ausencia podrían incluso considerarse atípicos. Por esta razón se considera realizar un diagrama de cajas (*boxplot*):



Acá vemos visualmente la asimetría y se identifican con claridad algunos *outliers*, estando el 75% de los datos de ausencia por debajo de las 10 horas. Adicionalmente, se observa que hay una asimetría en la distribución según la posición de la media en la caja y en la separación de los “bigotes”.

También, a partir de los plots anteriores se decidió realizar la revisión de los registros donde la ausencia es cero (*no-ausencia*), esto es cerca del 6% de los datos, que corresponden a 44 registros.

```
## [1] 44
## [1] 5.945946
```

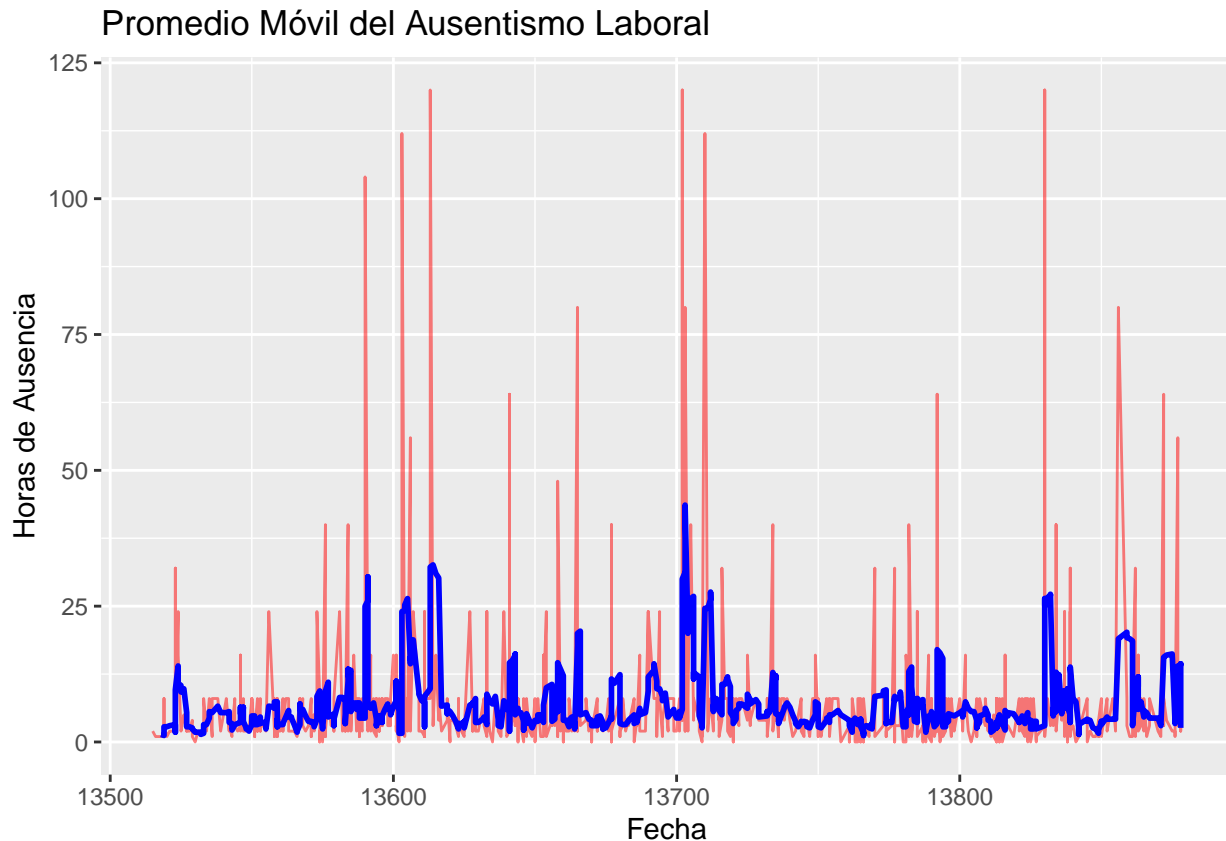
Por otro lado, el dataset no incluye una fecha específica por lo que es necesario construirla a partir de variables tales como `Month.of.absence` y `Day.of.the.week`. Para ello, usamos el siguiente código.

Media Móvil

El ausentismo laboral día a día puede tener mucha variabilidad, pues algunos días hay cero horas y otros días hay ausencias largas. La media móvil reduce las fluctuaciones diarias y muestra una tendencia más limpia. En este caso se generó un gráfico que muestra en rojo la variabilidad real y en azul la media móvil.

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_line()`).

## Warning: Removed 7 rows containing missing values or values outside the scale range
## (`geom_line()`).
```



Se aplicó un promedio móvil de 5 días al ausentismo laboral registrado. A partir del gráfico se puede concluir que en el primer trimestre el ausentismo suele ser bajo a comparación de los siguientes trimestres, sobre todo el tercer que presenta un pico promedio de más de 50 horas

Rezagos

El análisis de rezagos busca identificar si el ausentismo de hoy está relacionado con el ausentismo de días anteriores. Por ejemplo, interesa saber si un empleado que falta hoy tiene más probabilidad de faltar también mañana, o si, después de un día de alta ausencia, el ausentismo tiende a aumentar o disminuir en los días siguientes. Este tipo de relación temporal es fundamental para la construcción de modelos predictivos, como los modelos ARIMA o las regresiones que incorporan dependencias en el tiempo.

```
## [1] -0.05484861
```

```
## [1] 0.0468678
```

```
## [1] -0.02001964
```

El análisis de los rezagos a 1, 7 y 15 días muestra correlaciones muy bajas (0.0102, 0.0029 y 0.0417 respectivamente), indicando la ausencia de dependencia temporal significativa a corto, mediano y largo plazo semanal en los datos de ausentismo laboral. Estos resultados sugieren que el ausentismo no sigue un patrón repetitivo ni está influenciado por ausencias anteriores, comportándose de manera aleatoria a lo largo del tiempo. Esto claramente está también influenciado por la forma como se construyó la serie de tiempo de forma aleatoria a partir de la información disponible.

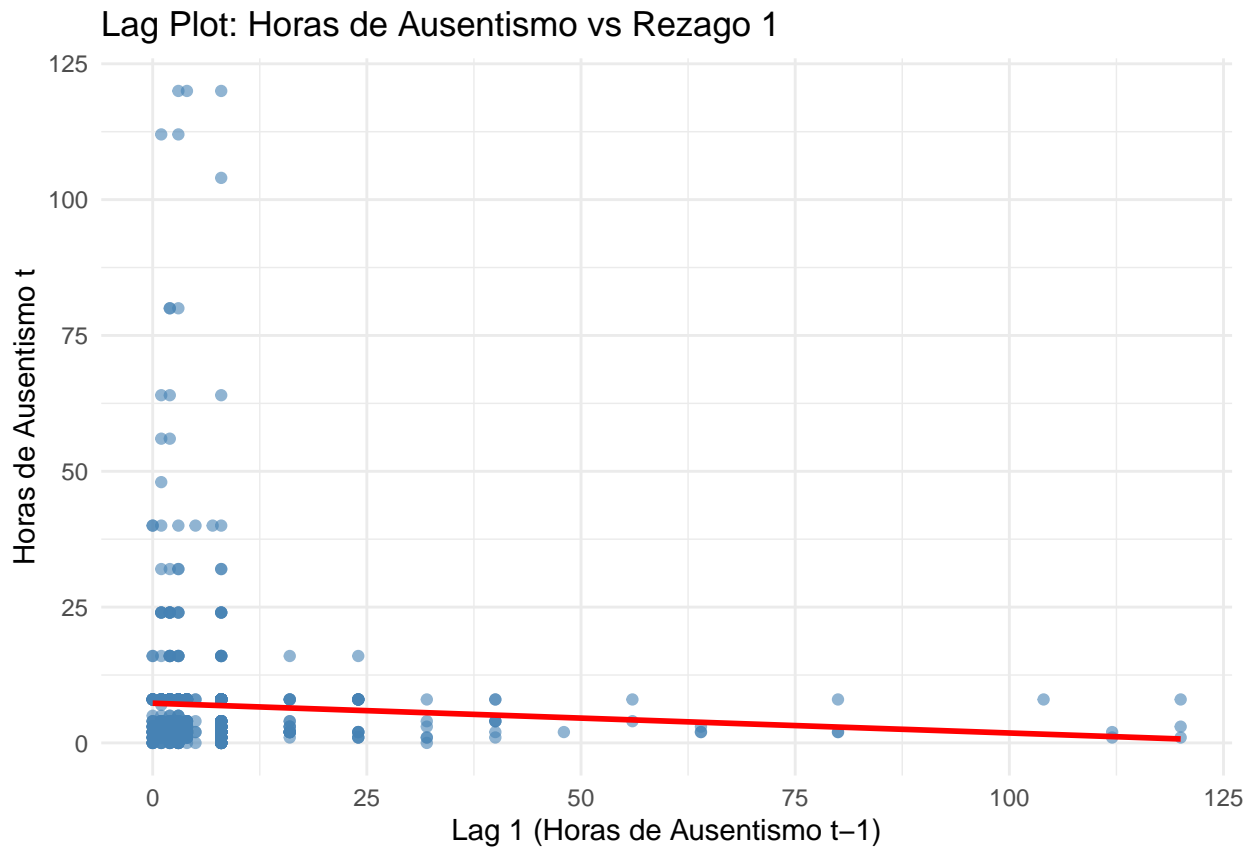
Este tipo de gráfico muestra la relación entre el valor en t y su valor en $t-1$, esto es:

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
```

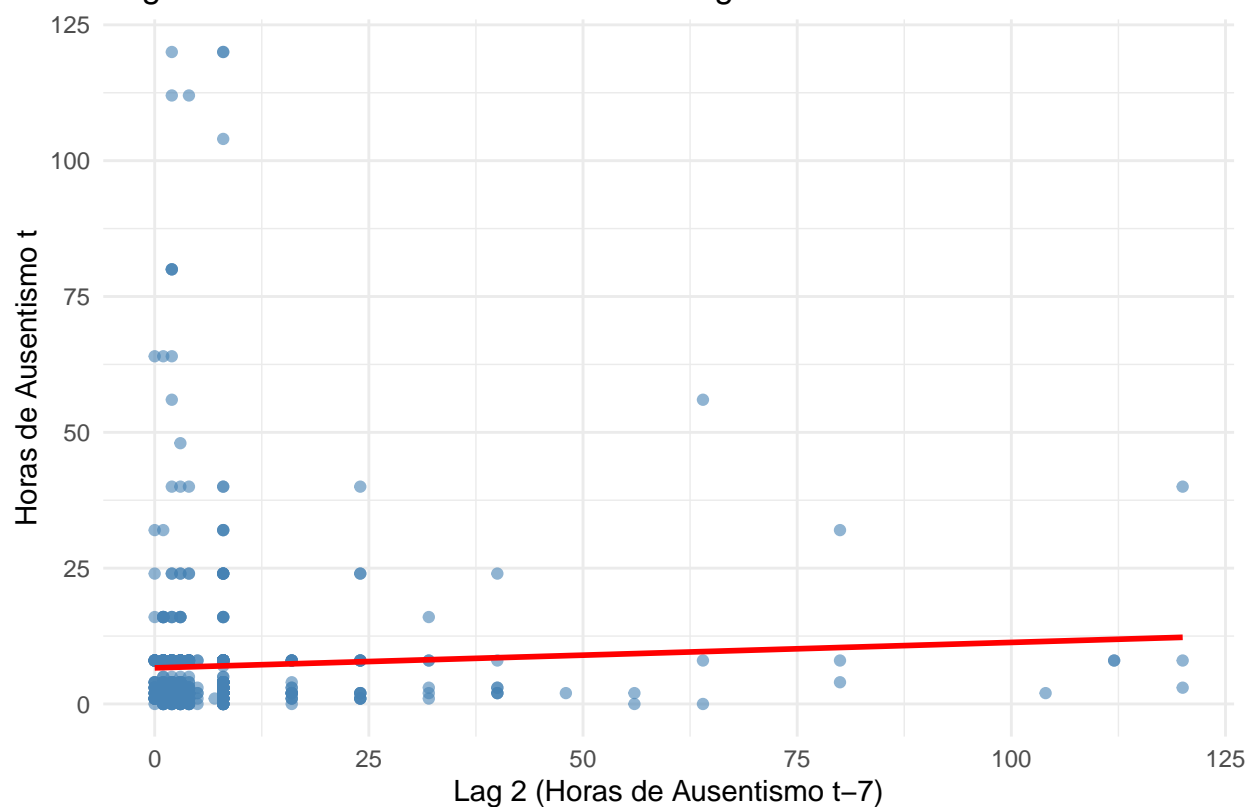
```
## (`stat_smooth()`).
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
```

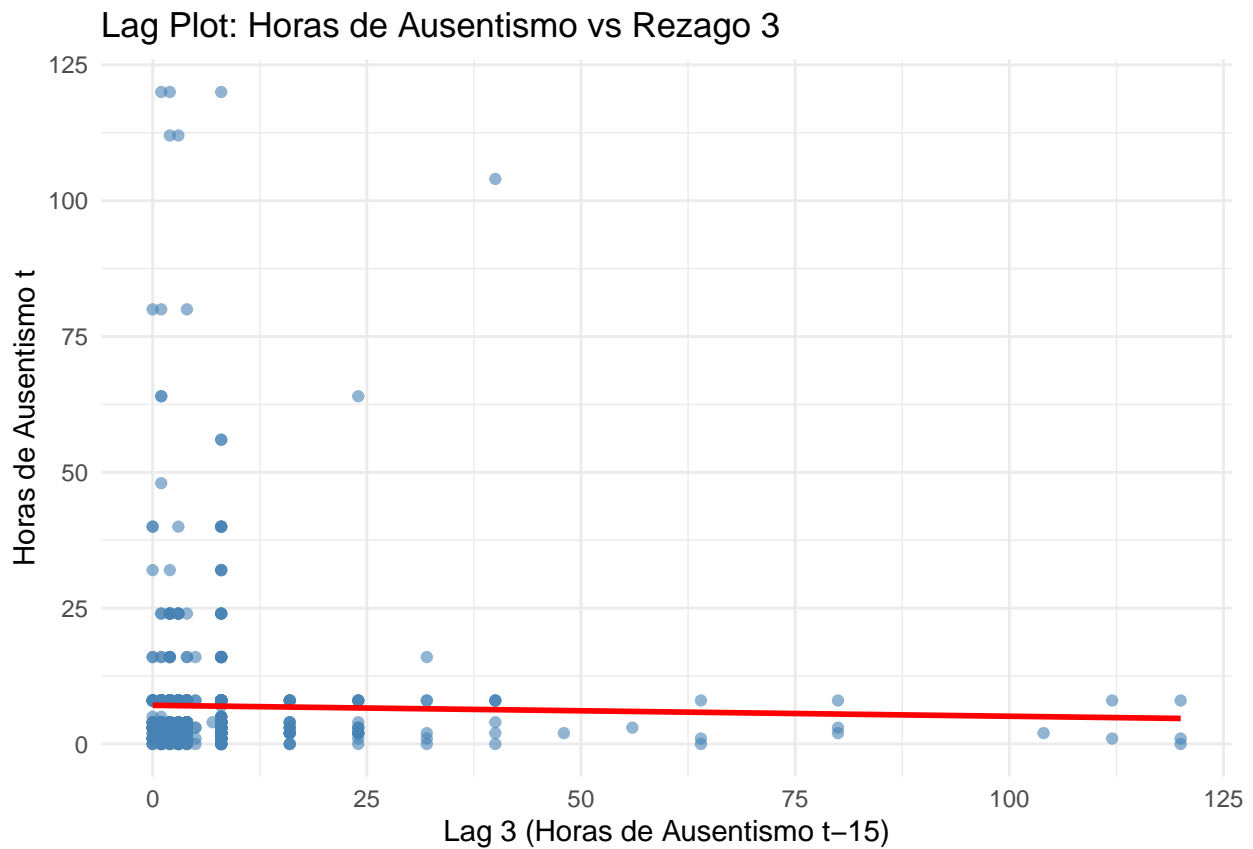


```
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 7 rows containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 7 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

Lag Plot: Horas de Ausentismo vs Rezago 2



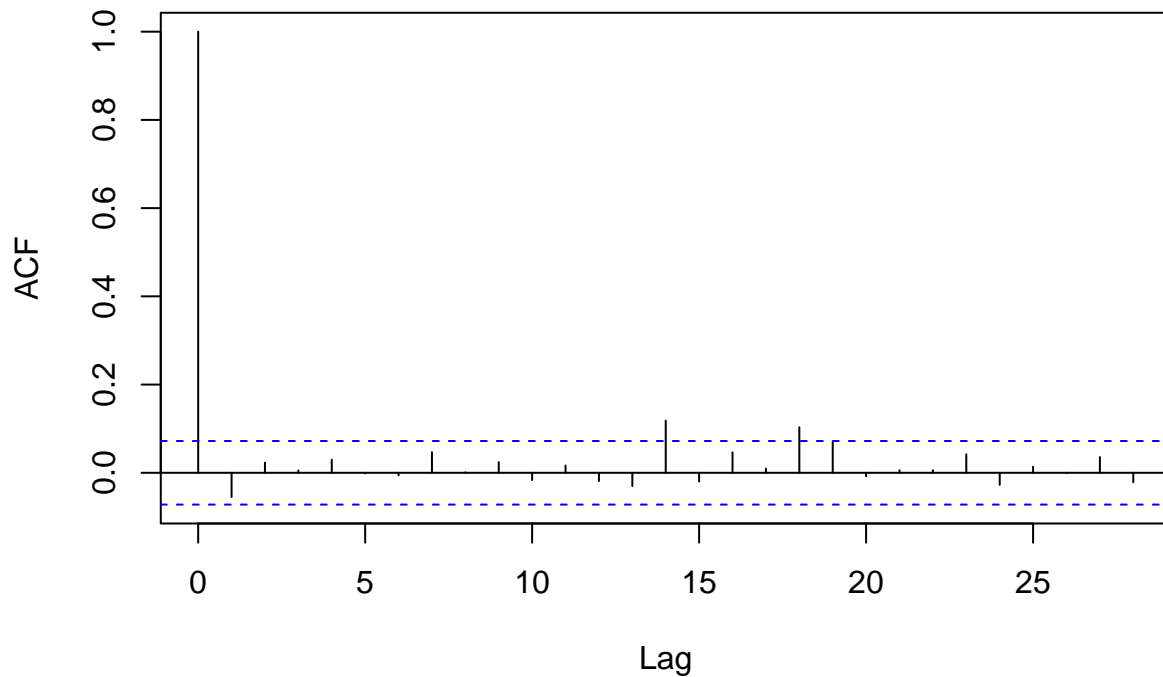
```
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 15 rows containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 15 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



En estos gráficos podemos observar que los puntos son caóticos o dispersos en todos los rezagos, no hay dependencia fuerte en el rezago, es decir no hay mucha relación entre el valor actual y el pasado.

vamos a graficar la *Autocorrelation Function (ACF)*, esto mide cuánta correlación existe entre la serie y sus propios rezagos.

ACF – Autocorrelación de ausencias



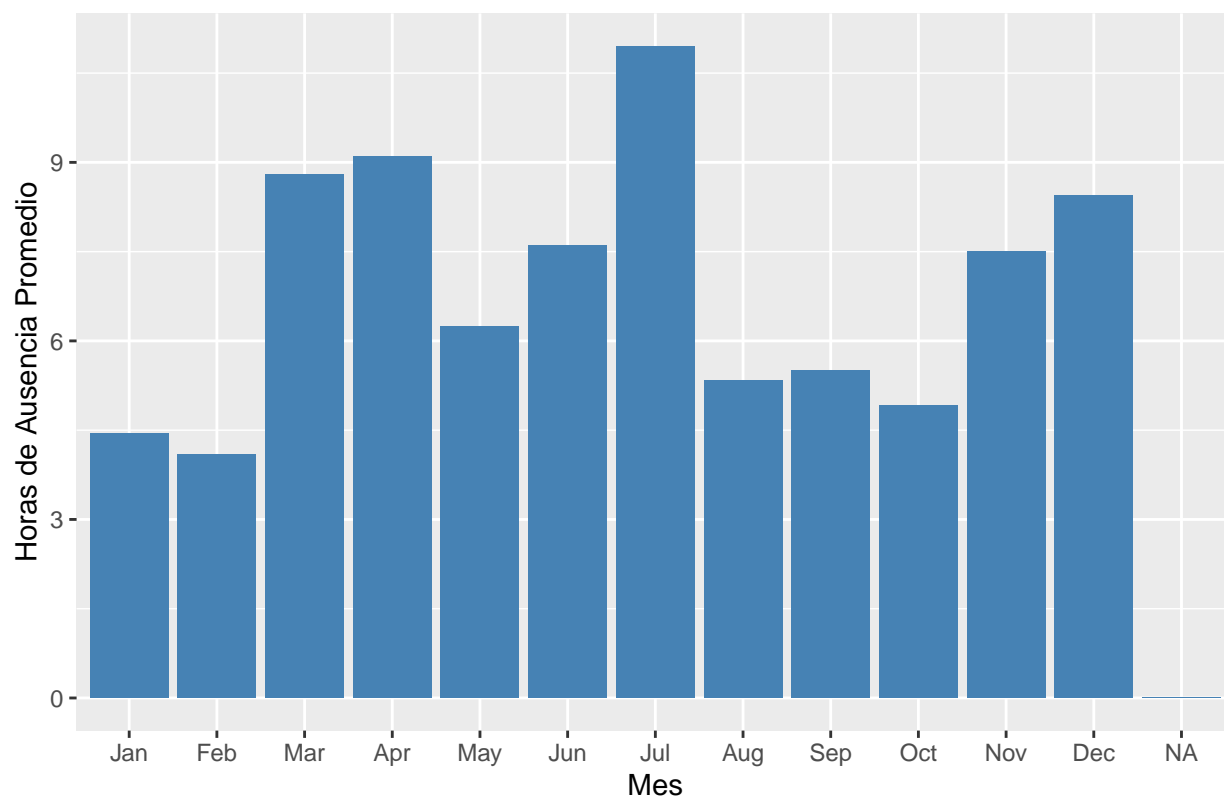
A partir del resultado del gráfico se interpreta que no hay correlaciones significativas en los rezagos.

Análisis de estacionalidad

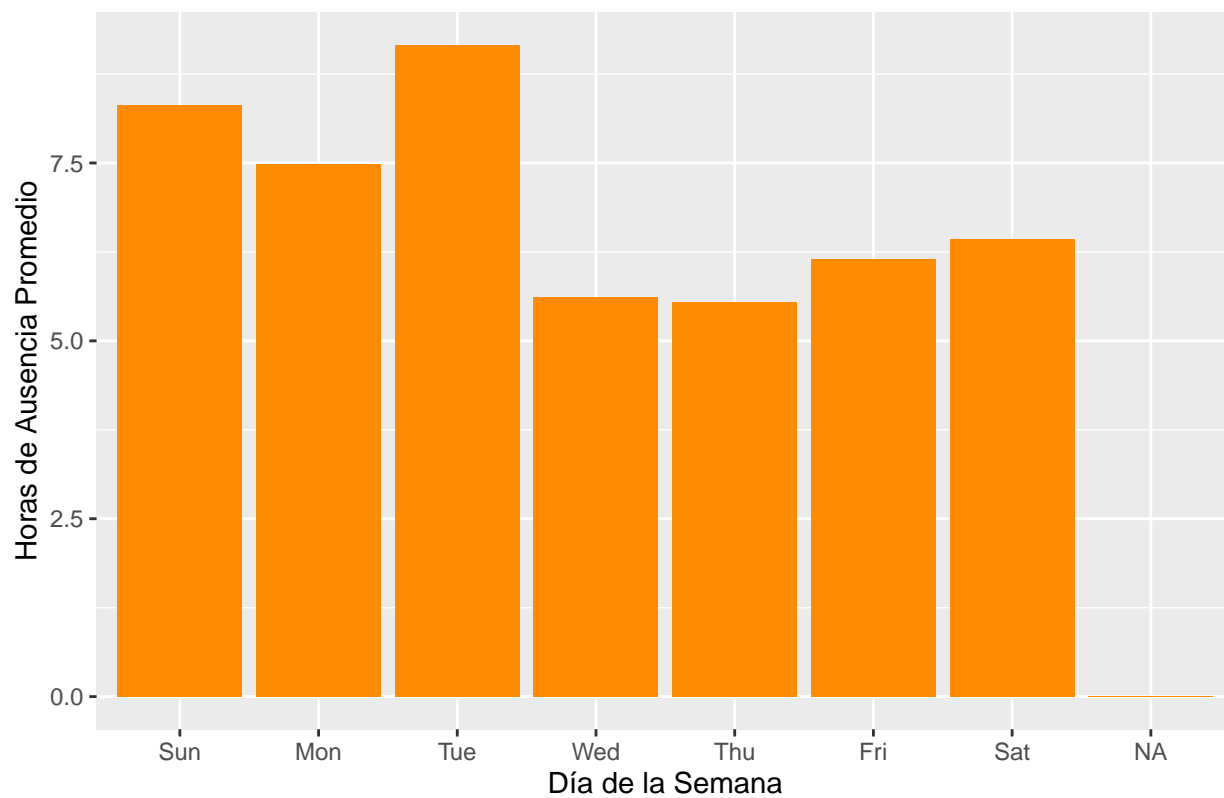
La estacionalidad es un patrón repetitivo o cíclico que ocurre a intervalos regulares en una serie de tiempo, como cada semana, cada mes o cada año.

En este caso, queremos ver si hay momentos del año (o del mes) en los que el ausentismo tiende a aumentar o disminuir sistemáticamente. Como en el conjunto de datos tenemos el mes (simulado con fechas) y las horas de ausentismo, podemos ver si ciertos meses muestran mayores ausencias que otros.

Estacionalidad Mensual del Ausentismo Laboral



Estacionalidad por Día de la Semana



El análisis de la estacionalidad mostró que el ausentismo laboral no presenta variaciones cíclicas fuertes a lo largo del año. Aunque se observan ligeras variaciones en algunos meses, no se identificaron patrones claros que indiquen un comportamiento estacional robusto. Esto sugiere que el ausentismo registrado responde más a factores puntuales que a eventos recurrentes en el tiempo, lo cual limita la capacidad predictiva basada en componentes estacionales.