

Calidad del agua superficial en el territorio mexicano

Jimena Pedroza & Tamara Guerrero

La calidad del agua superficial en el territorio mexicano es un tema de gran relevancia, ya que este recurso representa una de las principales fuentes de abastecimiento hídrico para el país. El agua superficial se encuentra en cuerpos como ríos, lagos, lagunas y presas, y se utiliza ampliamente para consumo humano, riego agrícola, procesos industriales, generación de energía y conservación de ecosistemas. En México, aproximadamente el 61.3% del agua utilizada proviene de fuentes superficiales.

Este tipo de agua se origina principalmente a partir del escurrimiento de la lluvia sobre la superficie terrestre, la cual se concentra en cuencas hidrográficas distribuidas en todo el territorio nacional. A diferencia del agua subterránea, el agua superficial está en constante movimiento, lo que facilita su renovación natural, pero también la hace más vulnerable a la contaminación. Diversos factores como el crecimiento urbano, la actividad industrial, la agricultura intensiva y el cambio climático han deteriorado la calidad del agua en muchas regiones del país.

La Comisión Nacional del Agua (CONAGUA), a través de la Red Nacional de Medición de la Calidad del Agua (RENAMECA), realiza un monitoreo constante del estado de los cuerpos de agua superficial. Este monitoreo se basa en una serie de indicadores fisicoquímicos y biológicos que permiten evaluar su condición. Entre los principales indicadores se encuentran la Demanda Bioquímica de Oxígeno (DBO₅), que mide la cantidad de oxígeno requerido para descomponer la materia orgánica biodegradable, donde valores superiores a 30 mg/L indican una contaminación severa; el Oxígeno Disuelto (OD), esencial para la vida acuática, cuyo nivel óptimo es mayor a 5 mg/L; los Sólidos Suspendidos Totales (SST), que reflejan la turbidez del agua; y los niveles de nutrientes como nitratos y fosfatos, cuya acumulación en exceso puede provocar eutrofización, un fenómeno que reduce el oxígeno disponible y altera los ecosistemas.

También se monitorean los coliformes fecales, los cuales indican contaminación biológica de origen humano o animal, y los metales pesados como plomo, mercurio y arsénico, que representan un riesgo para la salud cuando superan los límites establecidos por las normas ambientales. Estos indicadores permiten tener una visión cuantitativa y objetiva sobre la calidad del agua superficial en diferentes regiones del país.

Según el informe más reciente publicado por la CONAGUA en 2023, del total de cuerpos de agua superficial monitoreados, el 29% presenta una calidad considerada buena, el 35% tiene una calidad aceptable o moderada, y el 36% restante presenta una calidad mala o contaminada. Las zonas más afectadas se ubican en regiones densamente pobladas o con fuerte actividad industrial, como el Valle de México, la zona metropolitana de Guadalajara y partes del norte del país, donde la actividad agrícola intensiva también contribuye a la contaminación del agua.

La conservación del agua superficial es esencial para garantizar la salud de las personas, la viabilidad de las actividades productivas y la estabilidad de los ecosistemas acuáticos. Debido a su alta exposición a fuentes de contaminación y a su

naturaleza dinámica, es fundamental mantener un monitoreo constante, aplicar medidas de saneamiento efectivas y promover un uso responsable del recurso. La información proporcionada por la CONAGUA y la RENAMECA constituye una herramienta clave para la toma de decisiones y el diseño de políticas públicas orientadas a la protección y recuperación de la calidad del agua superficial en México.

Los contaminantes superficiales son:

Código	Nombre completo	Que mide
DBO	Demanda Bioquímica de Oxígeno	Cantidad de oxígeno que necesitan los microorganismos para descomponer materia orgánica. Valores altos indican contaminación orgánica.
DQO	Demanda química de oxígeno	Cantidad de oxígeno requerida para oxidar químicamente la materia orgánica e inorgánica. Mide la contaminación total.
SST	Sólidos suspendidos totales	Partículas sólidas en suspensión en el agua (arena, limo, materia orgánica). Valores altos afectan la claridad y calidad del agua.
CF	Coliformes fecales	Bacterias indicadoras de contaminación fecal. Su presencia sugiere riesgo sanitario.
E_COLI	Escherichia coli	Bacteria fecal específica. Presencia indica contaminación reciente y riesgo de enfermedades.
ENT_FE C	Enterococos fecales	Otro grupo de bacterias fecales. Usadas como indicador de contaminación en agua recreativa.
OD%L	Oxígeno disuelto (porcentaje de saturación)	Porcentaje de oxígeno disponible para los organismos acuáticos. Bajos niveles indican deterioro ambiental.

TOX_L	Toxicidad letal	Indicador de toxicidad aguda para organismos de prueba. Puede deberse a metales pesados, pesticidas, etc.
-------	-----------------	---

También vamos a utilizar un Semáforo de la calidad del agua, que es una clasificación de la calidad, basada en intervalos de concentración de cada contaminante, por lo que se establece una escala de cumplimiento que asigna un color: verde que representa la mejor calidad, amarillo calidad intermedia y rojo mala calidad.

Nuestras preguntas de investigación son:

1. ¿Existen diferencias significativas en la calidad bacteriológica del agua entre municipios de Tlaxcala?
2. ¿Qué tan representativos son los promedios de DBO, DQO y SST para evaluar la calidad del agua superficial en Tlaxcala, considerando su variabilidad entre municipios?
3. ¿Cuáles son los factores ambientales y sociales que hacen que estos acuíferos estén más contaminados?

- Dashboard

El siguiente código mostrará los indicadores clave en la parte superior del dashboard:

```
### Indicadores generales

```{r}
#Total de sitios
valueBox(nrow(df_superficial), icon = "fa-tint", color = "lightblue", caption = "Sitios monitoreados")
```

450

```{r}
#Total de cuencas
valueBox(value = length(unique(df_superficial$CUENCA)), icon = "fa-mountain", color = "lightgreen", caption = "Cuencas
monitoreadas")
```

87
```

```

Column
-----

```{r}
Porcentaje de sitios con semáforo rojo
porc_rojo <- round(mean(df_superficial$SEMAFORO == "Rojo") * 100)
valueBox(porc_rojo, icon = "fa-exclamation-triangle", color = "red", caption = "% Semáforo Rojo")

Porcentaje de sitios con semáforo amarillo
porc_amarillo <- round(mean(df_superficial$SEMAFORO == "Amarillo") * 100)
valueBox(porc_amarillo, icon = "fa-exclamation-circle", color = "orange", caption = "% Semáforo Amarillo")

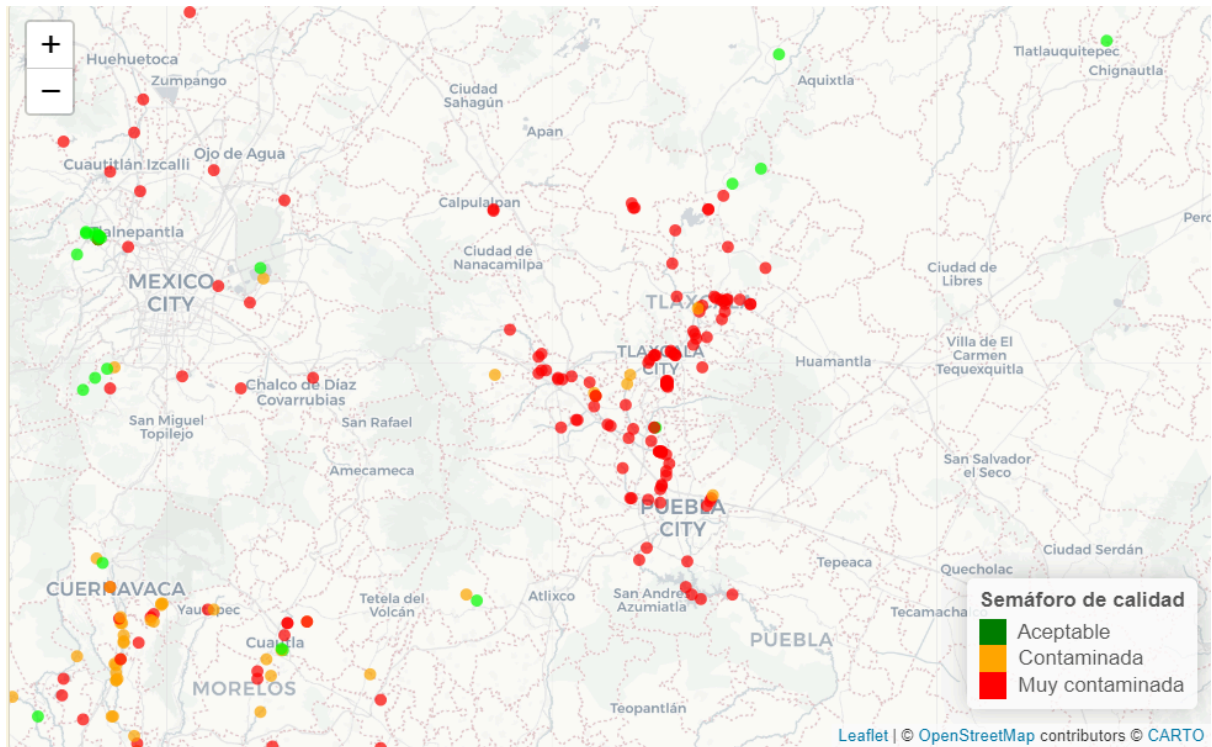
Porcentaje de sitios con semáforo verde
porc_verde <- round(mean(df_superficial$SEMAFORO == "Verde") * 100)
valueBox(porc_verde, icon = "fa-check-circle", color = "green", caption = "% Semáforo Verde")
```

```

Mapa interactivo:

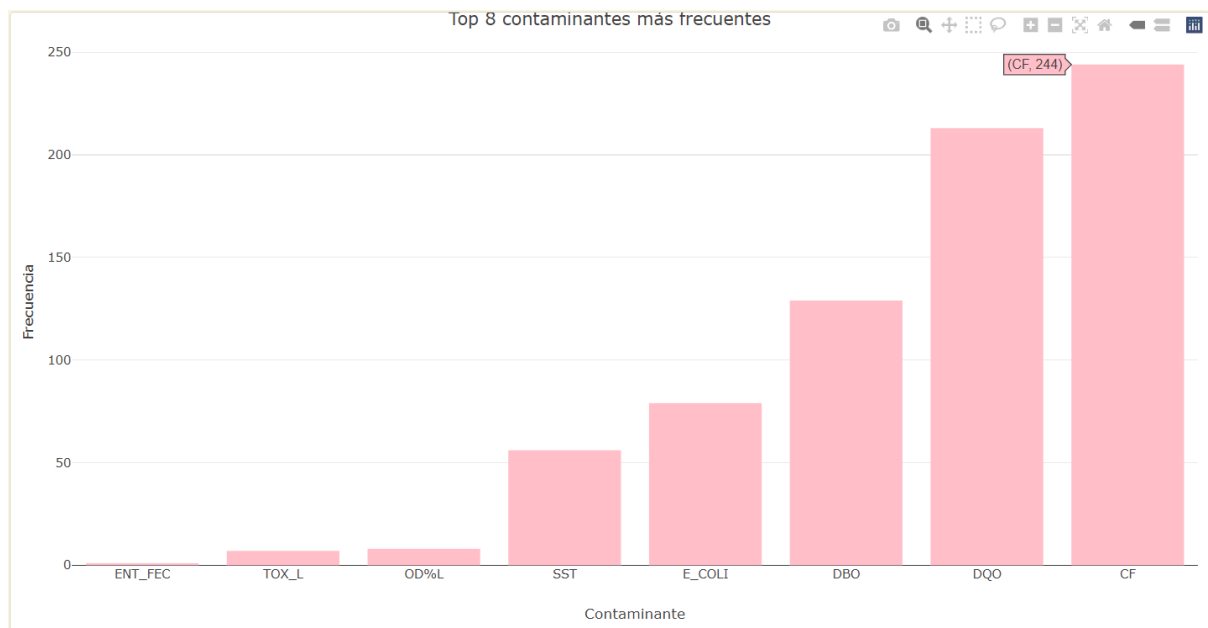


- Muestra todos los sitios sobre un mapa de fondo claro.
- Usa círculos codificados por color del semáforo.
- Muestra un popup con nombre del sitio y estado.

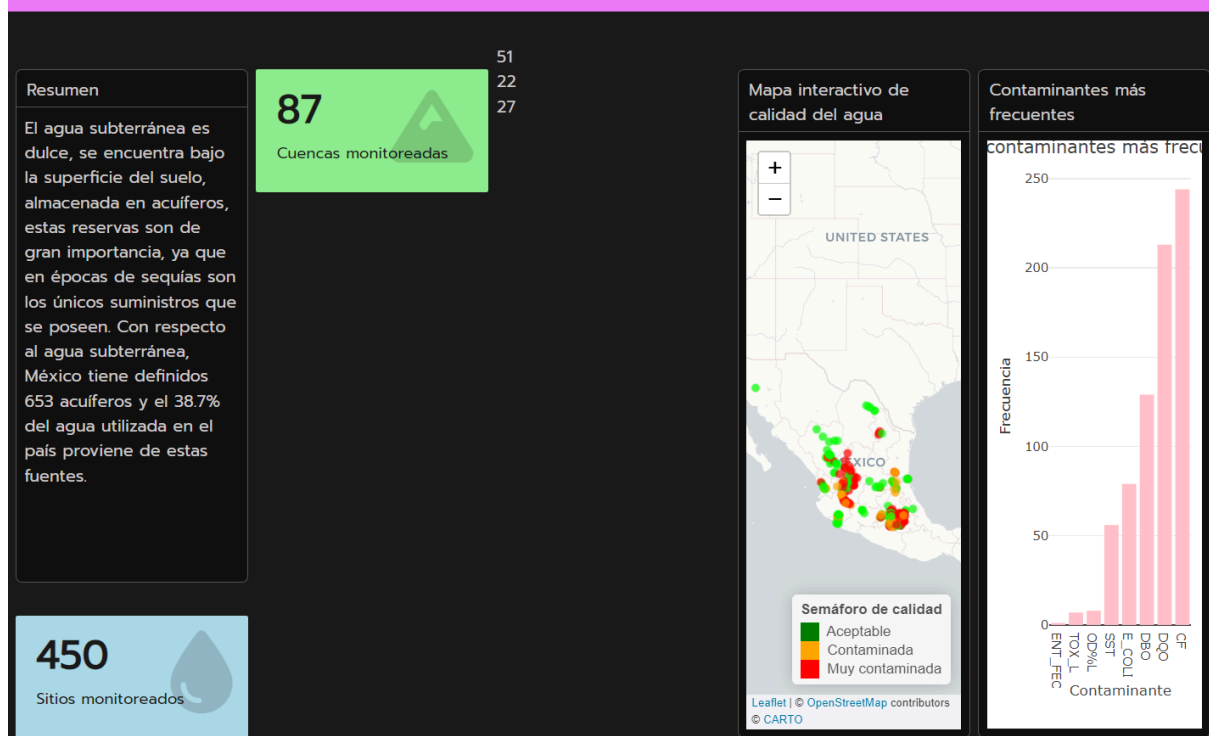


Aquí nos damos cuenta que Tlaxcala tiene muchos puntos rojos, esto nos da pie a que se haga un modelo para ver qué contaminantes son los que más afectan esta región de México.

Gráfica de barras interactiva



Resultado del dashboard



- Análisis ANOVA

En este caso la variable dependiente (lo que estamos midiendo), es el parámetro de calidad del agua que estamos evaluando, osea: DBO_mg/L, DQO_mg/L, SST_mg/L y COLI_FEC_NMP_100mL. Estas son las variables numéricas que cambian dependiendo del municipio.

Nuestra variable independiente (factor de agrupación), es la categoría por la que estamos comparando, en este caso: MUNICIPIO. Es la variable que define los grupos en el ANOVA. No cambia como resultado del análisis, sino que se usa para dividir los datos.

```

{r}
# ANOVA para DBO
anova_dbo <- aov(`DBO_mg/L` ~ MUNICIPIO, data = tlaxcala_filtrado)
summary(anova_dbo)

# ANOVA para DQO
anova_dqo <- aov(`DQO_mg/L` ~ MUNICIPIO, data = tlaxcala_filtrado)
summary(anova_dqo)

# ANOVA para SST
anova_sst <- aov(`SST_mg/L` ~ MUNICIPIO, data = tlaxcala_filtrado)
summary(anova_sst)

# ANOVA para Coliformes fecales
anova_coli <- aov(`COLI_FEC_NMP_100mL` ~ MUNICIPIO, data = tlaxcala_filtrado)
summary(anova_coli)

```

Resultado

```
      Df Sum Sq Mean Sq F value Pr(>F)
MUNICIPIO 24 335176    13966    0.74   0.78
Residuals 38 716865    18865
      Df Sum Sq Mean Sq F value Pr(>F)
MUNICIPIO 24 1133600    47233    0.774  0.743
Residuals 38 2318482    61013
      Df Sum Sq Mean Sq F value Pr(>F)
MUNICIPIO 24 259327    10805    1.288  0.238
Residuals 38 318821     8390
      Df Sum Sq Mean Sq F value Pr(>F)
MUNICIPIO 24 2.455e+15 1.023e+14    2.379 0.0082 **
Residuals 38 1.634e+15 4.301e+13
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Modelo Multinomial

```
# Cargar librerías
library(readxl)
library(dplyr)
library(nnet)
library(randomForest)
library(ggplot2)
|
# Leer la hoja de Excel
data <- read_excel("file_show.xlsx")
```

En esta parte del código preparamos y limpiamos los datos antes del análisis.

En este caso nos debemos asegurar que las variables que utilizamos están en el formato adecuado y de igual forma eliminar los datos incompletos.

```
...{r}
# Convertir semáforo en factor
data <- data %>%
  mutate(
    SEMAFORO = as.factor(SEMAFORO),
    contaminado = ifelse(SEMAFORO == "ROJO", 1, 0)
  ) %>%
  filter(!is.na(SEMAFORO) & !is.na(ESTADO) & !is.na(TIPO) & !is.na(CUENCA))

# Ver cuántos sitios hay por categoría
table(data$SEMAFORO)
...

```

También convertimos la columna SEMAFORO (que indica la calidad del agua: "ROJO", "VERDE", "AMARILLO") en un factor.

Los modelos matemáticos (como la regresión logística, multinomial o árboles de decisión) no entienden texto como "ROJO" o "VERDE". Por lo que se debe convertir internamente a variables numéricas.

Así R lo va a tratar como variable categórica, y no como texto plano.

Esto es esencial para modelos estadísticos y para gráficos categóricos.

contaminado = ifelse(SEMAFORO == "ROJO", 1, 0) Esta parte crea una nueva columna llamada contaminado.

Si el valor en SEMAFORO es "ROJO", le asigna 1 (es decir, sí está contaminado). Si no es "ROJO" (por ejemplo, "VERDE" o "AMARILLO"), le pone 0 (no está contaminado).

De esta forma se puede hacer un análisis binario: contaminado (1) vs. no contaminado (0).

filter(!is.na) Y en esta parte elimina las filas que tienen valores faltantes

multinom(): Es la función del paquete nnet que ajusta un modelo de regresión logística multinomial.

nuestra variable dependiente sería la categoría del Semáforo es decir lo que queremos predecir y la independiente sería Estado, Tipo, Cuenca.

El modelo busca entender cómo influye cada una de esas variables en la probabilidad de que el semáforo esté en ROJO, AMARILLO o VERDE.

```
library(nnet)

# Filtrar los datos solo para el estado de Tlaxcala
data_tlaxcala <- subset(data, ESTADO == "TLAXCALA")

# Multinomial logistic regression solo con datos de Tlaxcala
modelo_multinomial <- multinom(SEMAFORO ~ TIPO + CUENCA, data = data_tlaxcala)
summary(modelo_multinomial)

# Evaluar significancia
z <- summary(modelo_multinomial)$coefficients / summary(modelo_multinomial)$standard.errors
p_values <- (1 - pnorm(abs(z), 0, 1)) * 2
print(p_values)
```



```
# weights: 18 (10 variable)
initial value 86.790371
iter 10 value 30.182668
iter 20 value 29.908461
final value 29.907942
converged
Call:
multinom(formula = SEMAFORO ~ TIPO + CUENCA, data = data_tlaxcala)
```

```
Coefficients:
(Intercept) TIPOLÓTICO TIPOLÓTICO A - TIPO 1
ROJO      11.668075 -8.864745 -9.557327
VERDE      8.961021 -8.961996 -6.201279
TIPOLÓTICO A - TIPO 4 CUENCARÍO ALTO ATOYAC
ROJO      -8.353546 -1.012871
VERDE     -2.495750 -17.011807
```

```
Std. Errors:
(Intercept) TIPOLÓTICO TIPOLÓTICO A - TIPO 1
ROJO      88.23298  88.23599  88.24637
VERDE      88.23827  88.24393  587.50435
TIPOLÓTICO A - TIPO 4 CUENCARÍO ALTO ATOYAC
ROJO      88.24882  1.302246
VERDE     323.54163  307.288450
```

Residual Deviance: 59.81588

AIC: 79.81588

```
(Intercept) TIPOLÓTICO TIPOLÓTICO A - TIPO 1
ROJO      0.8947932  0.9199741  0.9137555
VERDE      0.9191100  0.9191065  0.9915783
TIPOLÓTICO A - TIPO 4 CUENCARÍO ALTO ATOYAC
ROJO      0.9245857  0.4366941
VERDE      0.9938453  0.9558508
```

```
library(dplyr)
library(ranger)

# Limpieza del texto
data <- data %>%
  mutate(SEMAFORO = toupper(trimws(SEMAFORO)))

# Filtrado para clasificación binaria
data_binaria <- data %>%
  filter(SEMAFORO %in% c("ROJO", "VERDE")) %>%
  mutate(
    contaminado = ifelse(SEMAFORO == "ROJO", 1, 0),
    contaminado = as.factor(contaminado),
    ESTADO = as.factor(ESTADO),
    TIPO = as.factor(TIPO),
    CUENCA = as.factor(CUENCA)
  )

# Entrenar modelo con ranger
modelo_rf <- ranger(
  contaminado ~ ESTADO + TIPO + CUENCA,
  data = data_binaria,
  importance = "impurity",
  probability = FALSE
)

# Imprimir resumen del modelo
print(modelo_rf)
```

```
# Importancia de variables
importance <- modelo_rf$variable.importance
print("Importancia de variables:")
print(sort(importance, decreasing = TRUE))
```

Ranger result

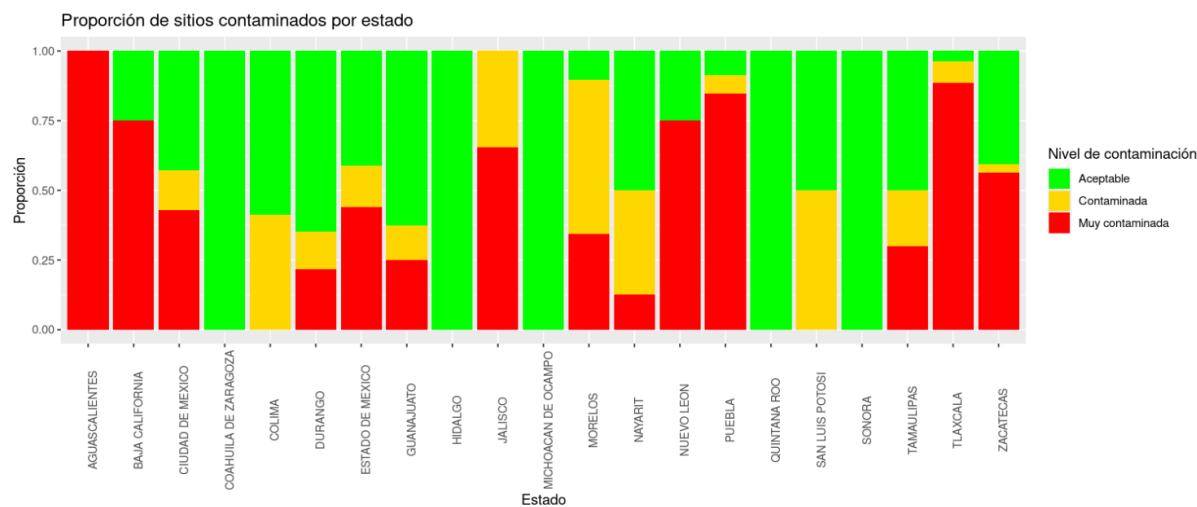
```
Call:
ranger(contaminado ~ ESTADO + TIPO + CUENCA, data = data_binaria, importance = "impurity", probability = FALSE)
```

```
Type: Classification
Number of trees: 500
Sample size: 350
Number of independent variables: 3
Mtry: 1
Target node size: 1
Variable importance mode: impurity
Splitrule: gini
OOB prediction error: 18.86 %
[1] "Importancia de variables:"
ESTADO CUENCA TIPO
40.97736 39.85648 26.21313
```

```
library(ggplot2)
library(dplyr)

# Asegúrate de que los niveles estén en el orden correcto
data <- data %>%
  mutate(SEMAFORO = factor(SEMAFORO, levels = c("VERDE", "AMARILLO", "ROJO")))

# Gráfico con colores personalizados y orden correcto
ggplot(data, aes(x = ESTADO, fill = SEMAFORO)) +
  geom_bar(position = "fill") +
  scale_fill_manual(
    values = c("VERDE" = "green", "AMARILLO" = "gold", "ROJO" = "red"),
    name = "Nivel de contaminación",
    labels = c("Aceptable", "Contaminada", "Muy contaminada")
  ) +
  labs(
    title = "Proporción de sitios contaminados por estado",
    y = "Proporción",
    x = "Estado"
  ) +
  theme(axis.text.x = element_text(angle = 90))
```



Conclusión

En nuestro análisis de ANOVA aplicados a los diferentes parámetros de calidad del agua por municipio en Tlaxcala cuando $\text{Pr}(> F)$: si es menor a 0.05, las diferencias entre municipios son estadísticamente significativas. El resultado indica que al menos un municipio tiene un valor significativamente distinto del resto para ese parámetro.

En el primer resultado (posiblemente DBO), el valor p ($\text{Pr}(> F)$) = 0.78 \rightarrow muy alto, entonces no hay diferencias estadísticamente significativas en los valores de DBO entre municipios. La variabilidad observada podría deberse al azar.

En el segundo resultado (probablemente DQO), el valor p = 0.743 \rightarrow muy alto, entonces no hay diferencias significativas en DQO entre municipios.

En el tercer resultado (probablemente SST), el valor p = 0.238 \rightarrow aunque más bajo, sigue siendo no significativo, la diferencia entre municipios no es estadísticamente significativa para SST.

En el último resultado (Coliformes fecales), el valor p = 0.0082 \rightarrow menor a 0.01, altamente significativo (**), por lo que sí hay diferencias significativas entre municipios en los niveles de coliformes fecales.

Respondiendo a nuestra primer pregunta de investigación, el ANOVA mostró una diferencia estadísticamente significativa para coliformes fecales (p = 0.0082). Nuestras variables eran: Municipio vs. Niveles de coliformes fecales, y nos dio como resultado que sí hay diferencias, algunos municipios presentan mayor contaminación fecal. Respondiendo a la segunda pregunta, aunque no hay diferencias significativas, puede haber variaciones extremas en algunos puntos de muestreo. Tal vez el promedio no es suficiente; se requieren medidas complementarias como rangos, desviación estándar o análisis espacial.

De acuerdo con el análisis de datos y el modelo de Random Forest, los factores que más contribuyen a la contaminación de cuerpos de agua superficial y que aplican como presión sobre acuíferos son principalmente la ubicación geográfica en donde las zonas con mayor actividad industrial y agrícola intensiva, como el Valle de México, el Alto Atoyac (Tlaxcala–Puebla) y partes del norte del país, muestran más sitios en condición roja. Esto se puede relacionar con mayor generación de residuos industriales, uso de fertilizantes y descargas sin tratamiento.

El tipo de cuerpo de agua ya que algunos tipos de cuerpos como presas o ríos en zonas urbanas están más expuestos a contaminantes, mientras que otros, como manantiales o arroyos menos accesibles, suelen mostrar mejores condiciones.

Esto responde a la pregunta de ¿Cuáles son los factores ambientales y sociales que hacen que las aguas superficiales estén más contaminadas?

Referencias

Dueñas, J. (2023, 12 septiembre). ¿Qué pasa con el agua de Tlaxcala? Enfermedad Renal.

<https://www.saludrenal.com.mx/tlx/2023/01/19/que-pasa-con-el-agua-de-tlaxcala/>

Del Agua, C. N. (s. f.). Calidad del agua en México

. . . gob.mx.

<https://www.gob.mx/conagua/articulos/calidad-del-agua>

De la Luz Degante, G. (2025, 23 marzo). Es más severo el problema de contaminación del agua que el de la escasez en Tlaxcala: Hernández; se prevé crisis hídrica en 2055. La Jornada de Oriente.

<https://www.lajornadadeoriente.com.mx/tlaxcala/es-mas-severo-el-problema-de-contaminacion-del-agua-que-el-de-la-escasez-en-tlaxcala-hernandez-se-preve-crisis-hidrica-en-2055/>