

# ENTREGA 1: GEOESTADÍSTICA

Santiago Vera Quiceno

Ricardo William Salazar Espinal

Daniel Felipe Puentes R.

ESTADISTICA ESPACIAL

2025-1

Francisco Javier Rodriguez



**Universidad Nacional de Colombia**

**Sede Medellín**

FACULTAD DE CIENCIAS

Medellín, Junio 2025

## Introducción:

En este estudio nos centraremos en el fenómeno de las altas presiones registradas en algunas regiones de Antioquia durante uno de los episodios de sequía más extremos de la zona: el mes de enero. En particular, tomaremos como caso de referencia el 15 de enero de 2025, en el intervalo entre las 12:00 m. y las 4:00 p. m., periodo en el que 354 municipios antioqueños se encontraron en alerta por riesgo de incendios forestales más info aquí.

El conjunto de datos de presión atmosférica horaria proviene de estaciones automáticas del IDEAM y entidades aliadas. Estos datos están disponibles en la plataforma datos.gov, constituyendo una base de datos muy interesante debido a la fiabilidad de sus mediciones y a que cubre de manera completa el territorio nacional.

*Nota: Debido a que hay varias mediciones de una misma estación, se va a realizar un promedio de las presiones por estación para el intervalo horario de interés.*

## Pregunta de Investigación:

¿Qué regiones de Antioquia, dentro del área cubierta por estaciones meteorológicas, presentan presiones atmosféricas más propicia para las altas temperaturas del 15 de enero, según su distribución espacial modelada mediante un Kriging Universal?

### 1. Parámetros y lectura de datos

Inicialmente, se procede a realizar la lectura de los datos de presión atmosférica, filtrando por el intervalo horario de interés (12:00 m. a 16:00 p. m.) y calculando el promedio de presión por estación.

```
##
## 12 13 14 15 16
## 162 166 156 158 168
```

### 2. Conversión a objetos espaciales

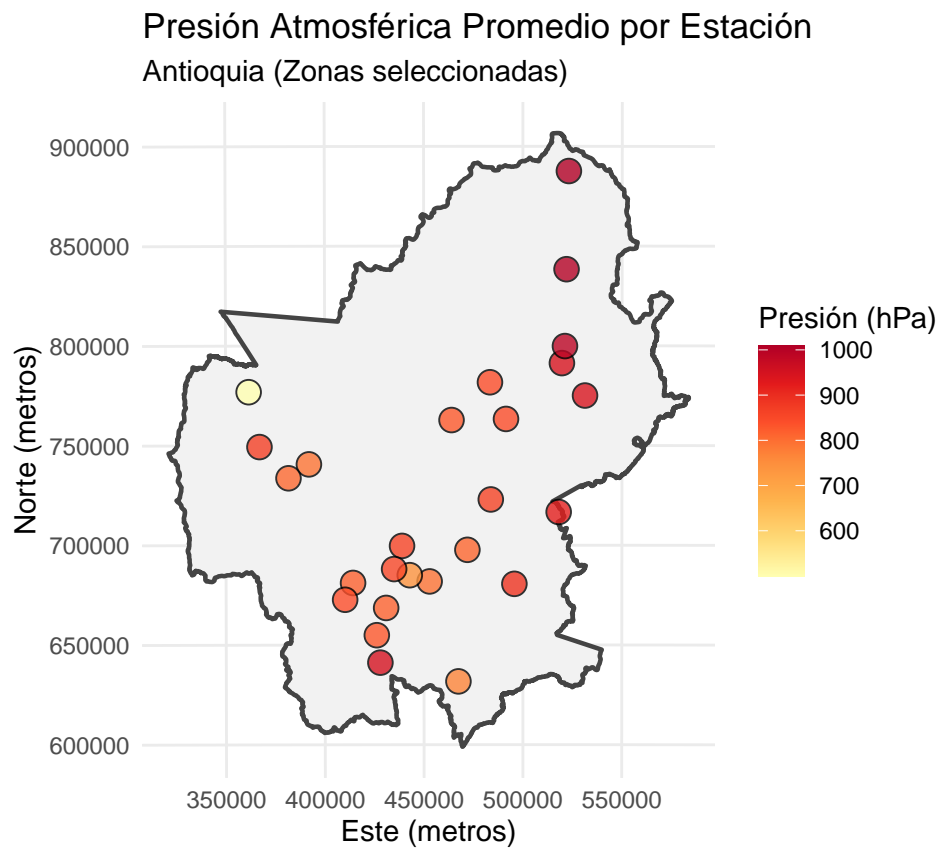
Una vez obtenido los datos de presión promedio por estación, se procede a convertirlos en un objeto espacial utilizando las coordenadas geográficas (Latitud y Longitud) y transformándolos al sistema de referencia UTM.

### 3. Análisis estructural de distancias

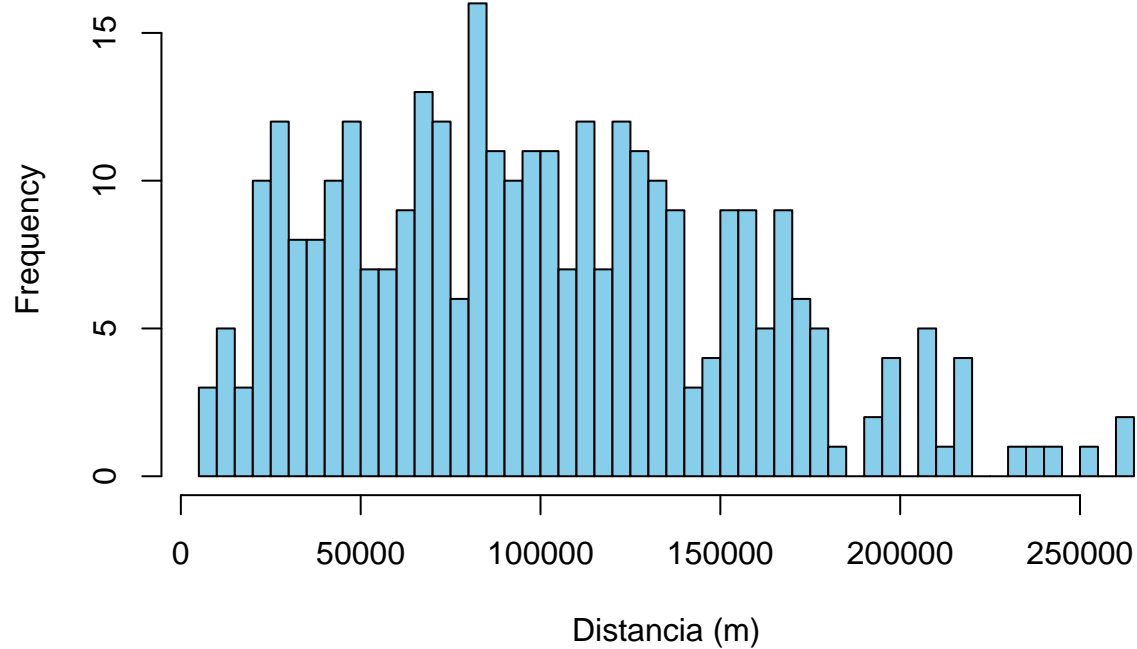
#### Datos Crudos

```
## # A tibble: 6 x 8
##   CodigoEstacion NombreEstacion Municipio ZonaHidrografica Latitud Longitud
##   <chr>          <chr>          <chr>    <chr>          <dbl>   <dbl>
## 1 0011110110     BOQUERON DE TOYO  GIRALDO  ATRATO - DARIÉN    6.70    -76.0
## 2 0011115040     MUSINGA          FRONTINO  ATRATO - DARIÉN    6.78    -76.2
## 3 0023085080     GRANJA EXPERIMENTA~ SAN ROQUE  MEDIO MAGDALENA    6.48    -74.8
## 4 0023085160     CORRIENTES       SAN VICE~  MEDIO MAGDALENA    6.31    -75.3
## 5 0023085220     SAN CARLOS       SAN CARL~  MEDIO MAGDALENA    6.16    -75.0
## 6 0023085270     AEROPUERTO J.M. CO~ RIONEGRO  MEDIO MAGDALENA    6.17    -75.4
## # i 2 more variables: valor_promedio <dbl>, n_obs <int>
```

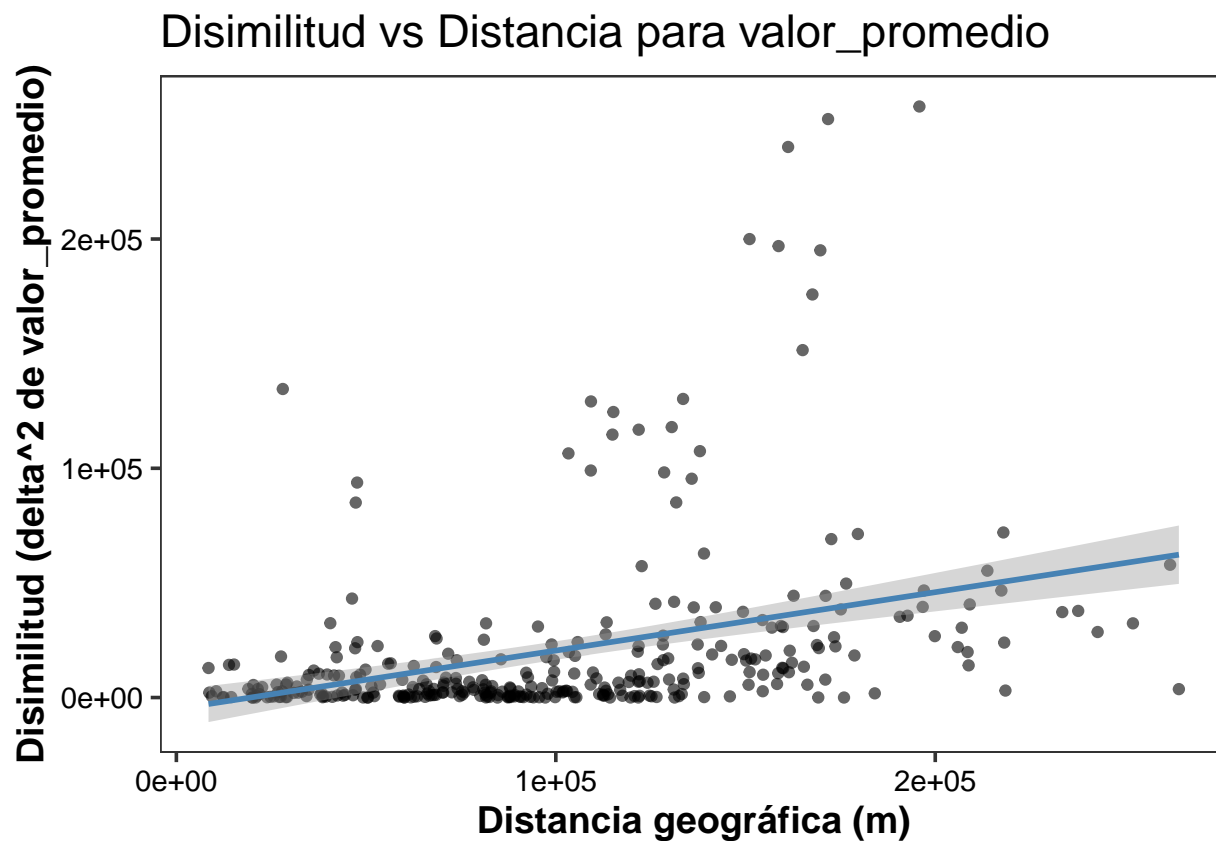
A continuación se realiza un análisis estructural de los datos crudos, visualizando la distribución de las presiones promedio por estación y generando un mapa de puntos usando leaflet.



## Distribución de distancias entre estaciones

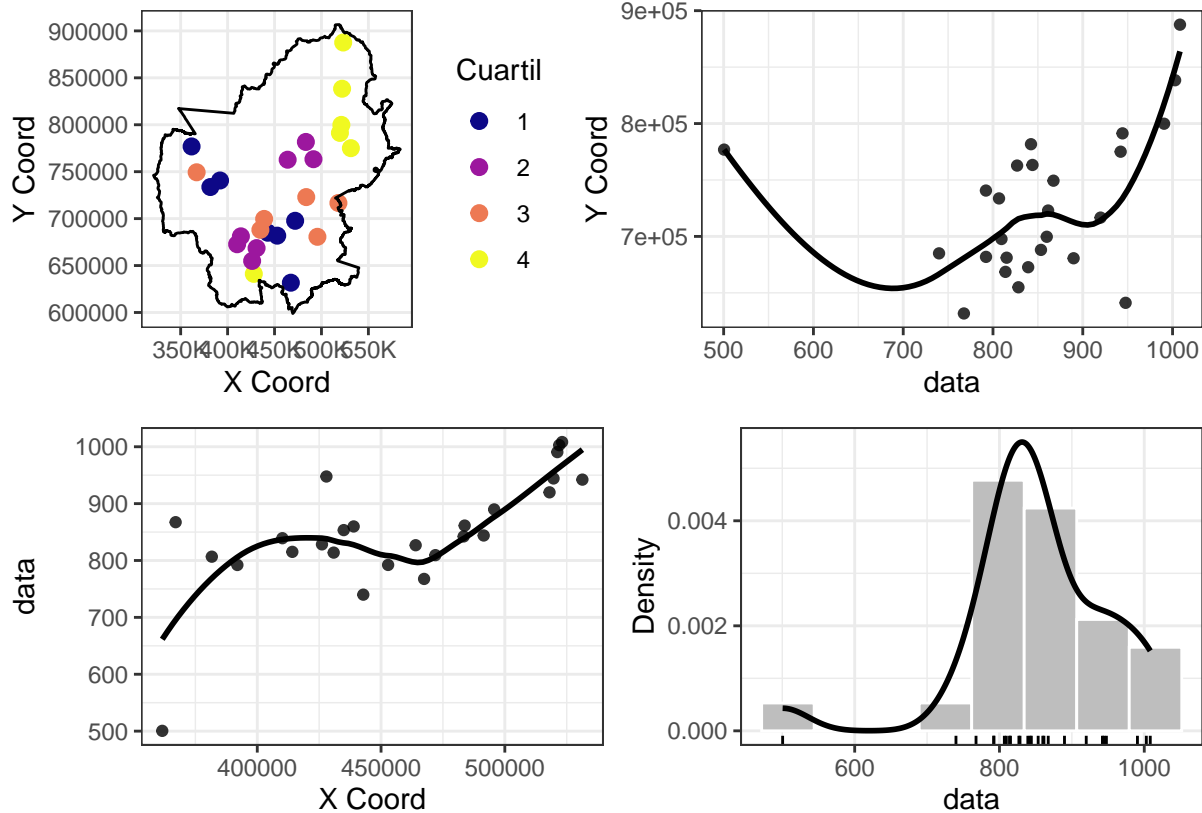


La distribución de distancias se observa que la mayoría de las distancias cae entre 20 y 130 km, con un pico cercano en los 80 km, también hay muy pocas distancias mayores a 200 km, lo que indica que las estaciones están relativamente cerca unas de otras, con una densidad de estaciones alta en el centro y noreste de Antioquia.



Para el variograma cloud, se observa una tendencia lineal positiva, lo que indica que a medida que aumenta la distancia geográfica entre las estaciones, también aumenta la disimilitud en los valores promedio de presión atmosférica. Esto sugiere una dependencia espacial entre las estaciones, donde las estaciones más cercanas tienden a tener valores más similares.

A continuación, se analiza el gráfico de tendencias y residuos:

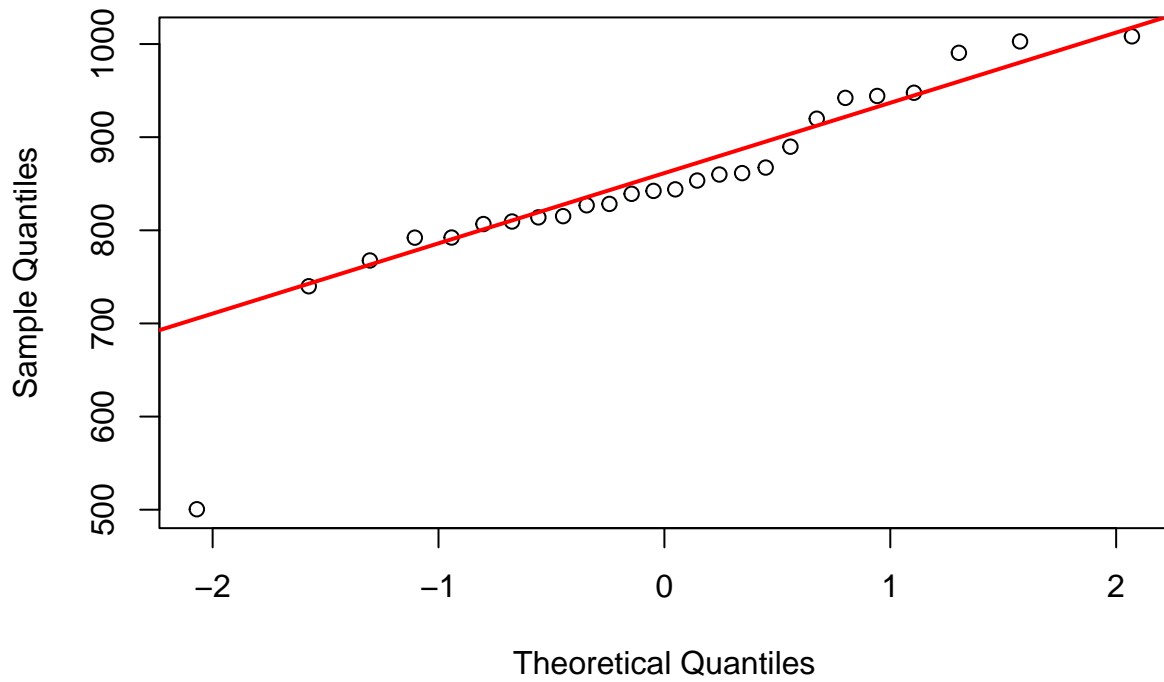


Los datos categorizados por cuantiles, muestran una fuerte estructura espacial. Se observa una clara tendencia con valores que aumentan de noroeste a sureste/noreste. Los gráficos de dispersión confirman la tendencia espacial observada; en el eje x los valores tienden a aumentar a medida que nos desplazamos hacia el este, mientras que para el eje y, los valores también tienden a aumentar a medida que nos desplazamos hacia el norte. La relación no es perfectamente lineal, lo que justifica la necesidad de modelar esta tendencia.

Por ultimo, el histograma indica que los datos no siguen una distribución normal.

```
##
##  Shapiro-Wilk normality test
##
## data:  z
## W = 0.87331, p-value = 0.004172
```

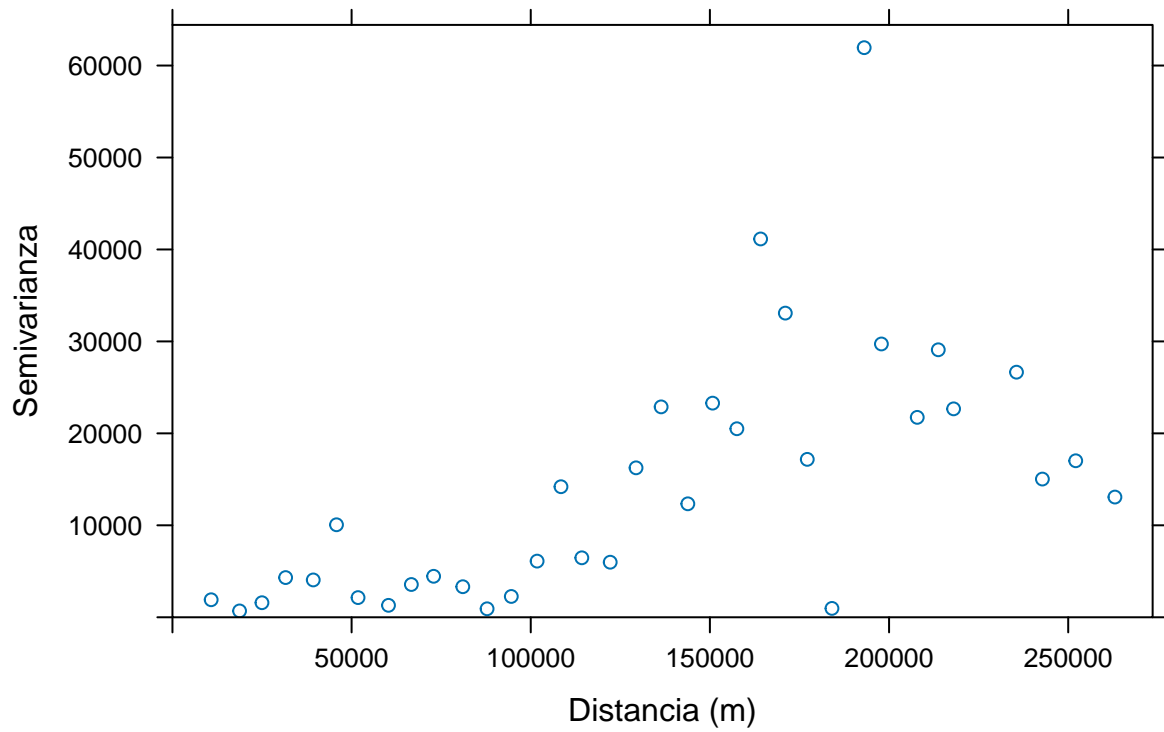
### QQ-plot de valor\_promedio (datos crudos)



Al realizar la prueba de normalidad de Shapiro-Wilk sobre los datos, este confirma lo ya mencionado en el gráfico anterior. Los datos de presión promedio no siguen una distribución normal, por lo cual se procede a usar una estimación clásica para el semivariograma puesto a que es menos restrictivo sobre los supuestos de normalidad.

#### 4. Semi-variograma empírico

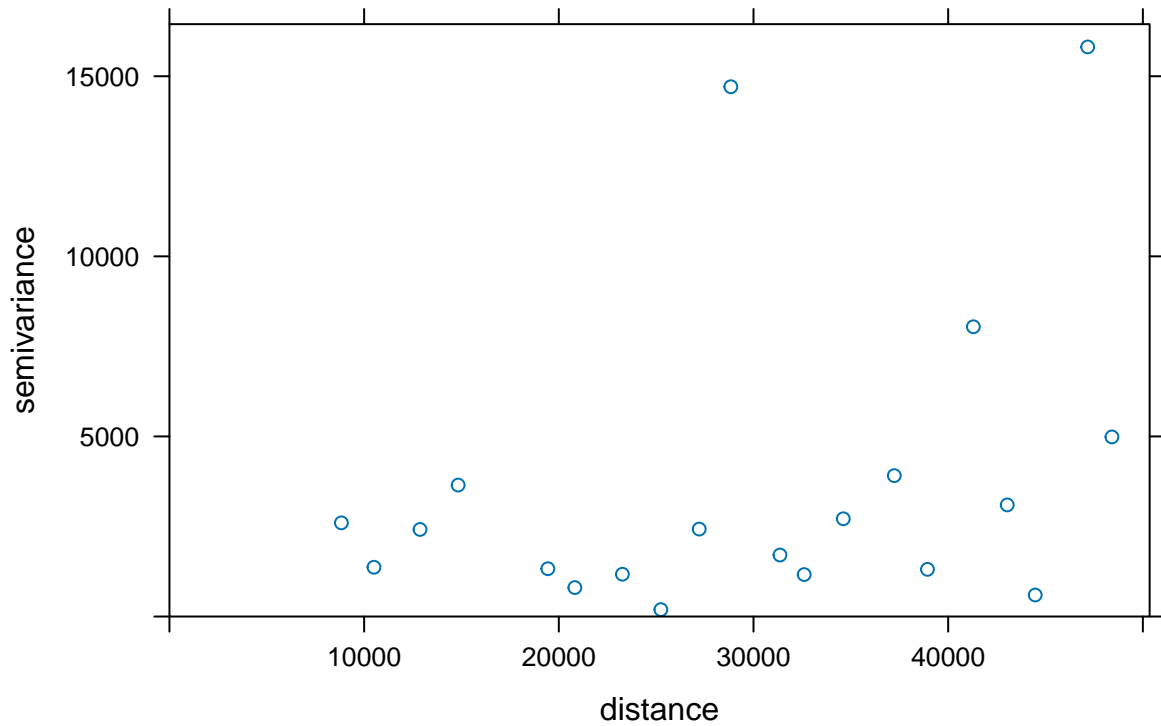
##### **Semi-variograma empírico de presiones promedio**



Se realiza un corte en el Semi-variograma para evitar el efecto de borde. En este caso, se establece un cutoff de 49 km y un ancho de 2 km.



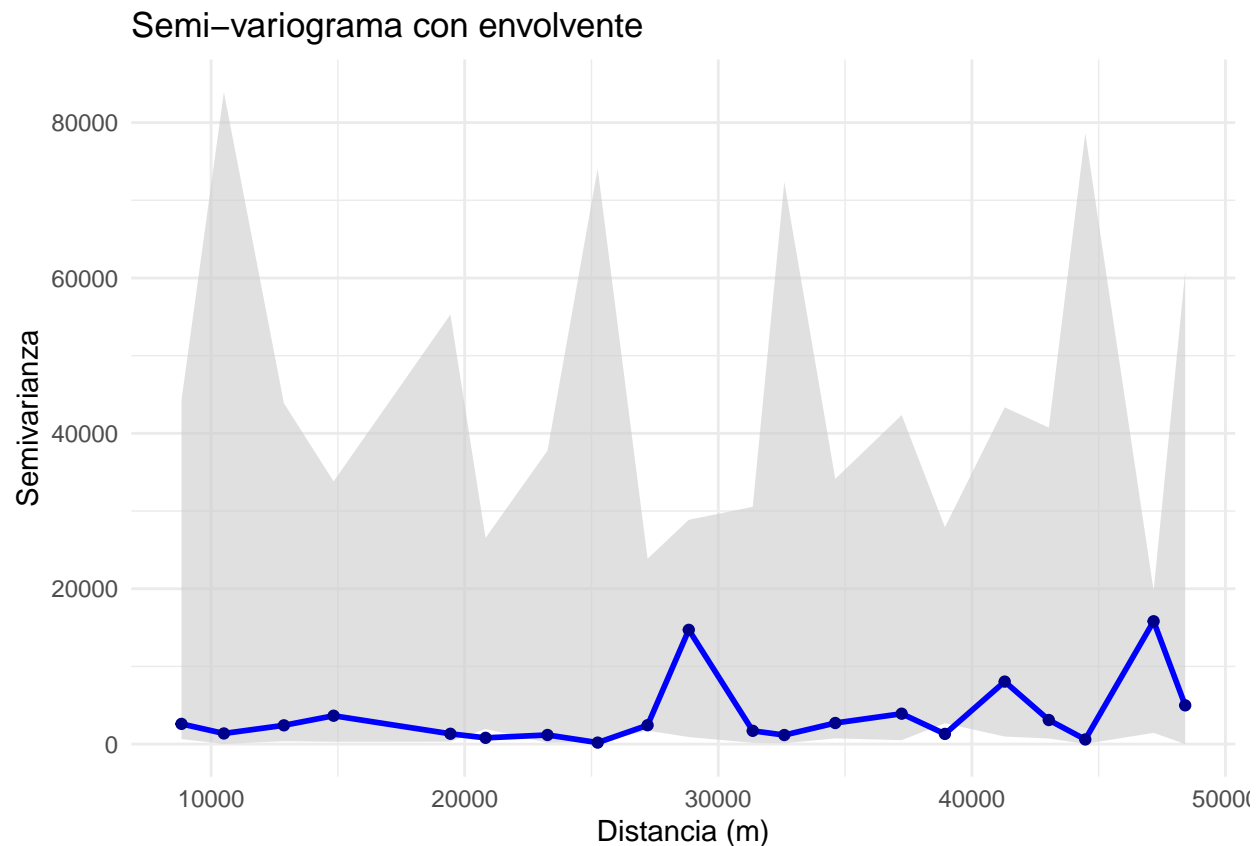
## Semi-variograma de presión promedio



Al analizar el semivariograma, este muestra un patrón creciente de la semivarianza con la distancia hasta un rango de 49 km, donde parece estabilizarse (Meseta~16.000), indicando que más allá de estas distancias la correlación espacial se pierde. La dispersión a mayores distancias sugiere un efecto de ruido o variabilidad no explicada por la distancia, mientras que el nugget inicial refleja variabilidad a escalas menores o errores de medición. El variograma con un corte en 49km reduce el ruido a largas distancias y es más adecuado para predicciones locales.

Cabe recalcar que usamos una distancia maxima de 49km debido al contexto del problema, a pesar de que más allá de la distancia mencionada se podría observar una meseta, el rango de 49km es suficiente para capturar la estructura espacial de los datos sin introducir ruido innecesario.

## 5. Envoltente del Semi-variograma por permutaciones



Al analizar el gráfico de la envoltente tenemos que mediante el uso del variograma estimado , junto con la envoltente obtenida por las 99 permutaciones, muestra que la curva azul supera claramente el contorno gris en varios tramos de distancia. Esto indica la presencia de estructura espacial significativa, ya que las semivarianzas observadas superan lo esperado bajo aleatoriedad. Este hallazgo, coherente con los análisis previos, será confirmado mediante el test de Mantel, el cual permitirá validar estadísticamente la dependencia espacial entre las distancias geográficas y las diferencias en los valores observados.

## 4. Test de Mantel

Finalmente, se realiza un test de Mantel para evaluar la relación entre las distancias geográficas y las disimilitudes de los valores promedio de presión atmosférica. Este test permite determinar si existe una correlación significativa entre las dos matrices de distancias.

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = dist_geo, ydis = dist_att, method = "pearson",      permutations = 999)
##
## Mantel statistic r: 0.4548
##      Significance: 0.001
##
## Upper quantiles of permutations (null model):
```

```
## 90% 95% 97.5% 99%
## 0.143 0.264 0.342 0.367
## Permutation: free
## Number of permutations: 999
```

Realizando 999 permutaciones con un valor de referencia de  $\alpha = 0.05$ , se obtiene un valor de p de 0.001, lo que indica una correlación significativa entre las distancias geográficas y las disimilitudes de los valores promedio de presión atmosférica. Esto sugiere que la presión atmosférica en Antioquia está influenciada por la distancia entre las estaciones de medición. También se puede apreciar que el estadístico de Mantel es de 0.45, lo que indica una correlación moderada entre las dos matrices de distancias, algo visto anteriormente en el gráfico del Semi-variograma cloud.

## 5. Estimación de parámetros

Para ajustar un modelo geoestadístico a los datos de presión atmosférica, se utilizará la librería `geoR` para realizar el ajuste de modelos de covarianza. Se emplearán dos métodos: Máxima Verosimilitud Restringida (REML) y Máxima Verosimilitud (ML).

```
## kappa not used for the gaussian correlation function
## -----
## likfit: likelihood maximisation using the function optim.
## likfit: Use control() to pass additional
##       arguments for the maximisation function.
##       For further details see documentation for optim.
## likfit: It is highly advisable to run this function several
##       times with different initial values for the parameters.
## likfit: WARNING: This step can be time demanding!
## -----
## likfit: end of numerical maximisation.

## kappa not used for the gaussian correlation function
## -----
## likfit: likelihood maximisation using the function optim.
## likfit: Use control() to pass additional
##       arguments for the maximisation function.
##       For further details see documentation for optim.
## likfit: It is highly advisable to run this function several
##       times with different initial values for the parameters.
## likfit: WARNING: This step can be time demanding!
## -----
## likfit: end of numerical maximisation.

## [1] "Resultados REML:"

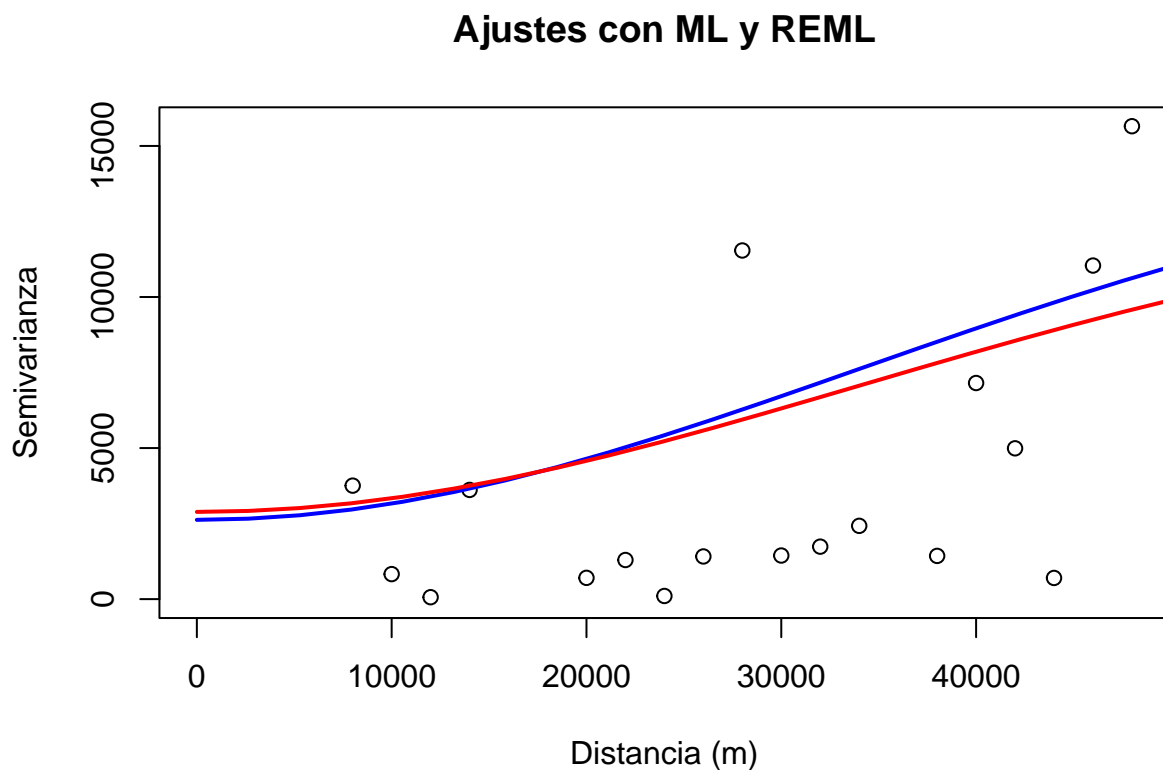
## likfit: estimated model parameters:
##      beta      tausq   sigmasq      phi
## " 856.4" " 2623.7" "12659.5" "48000.0"
## Practical Range with cor=0.05 for asymptotic range: 83079.28
##
## likfit: maximised log-likelihood = -147.3

## [1] "Resultados ML:"
```

```
## likfit: estimated model parameters:
##      beta      tausq   sigmasq      phi
## " 856.7" " 2886.7" "10590.0" "48000.0"
## Practical Range with cor=0.05 for asymptotic range: 83079.28
##
## likfit: maximised log-likelihood = -153.6
```

Para comparar los ajustes de los modelos REML y ML, se graficará el variograma empírico junto con las líneas de los modelos ajustados.

```
## variog: computing omnidirectional variogram
```



Ambos modelos se ajustan bien al variograma empírico. Para el análisis posterior, se seleccionaron los parámetros del método ML por sus propiedades estadísticas de menor sesgo.

Donde:

- **Rango:** Para este valor, se obtuvo un rango de 48000 metros, lo que indica la distancia a la cual la correlación espacial entre las estaciones de presión comienza a decrecer significativamente.
- **Nugget:** El valor del nugget es de 2886.7, lo que sugiere que hay una variabilidad a escalas menores que no está capturada por el modelo, posiblemente debido a errores de medición o variabilidad a escalas muy pequeñas.
- **Meseta:** La meseta del modelo es de 10590, lo que indica el valor de semivarianza a distancias mayores al rango, sugiriendo que la variabilidad espacial se estabiliza en este valor.

## 6. Predicción

Utilizando los parámetros del modelo ML, se realizó una interpretación espacial mediante Kriging Universal. Se eligió este método porque como se mencionó anteriormente en el análisis estructural, se detectó una tendencia en los datos, donde los valores de presión varían en función de sus coordenadas geográficas.

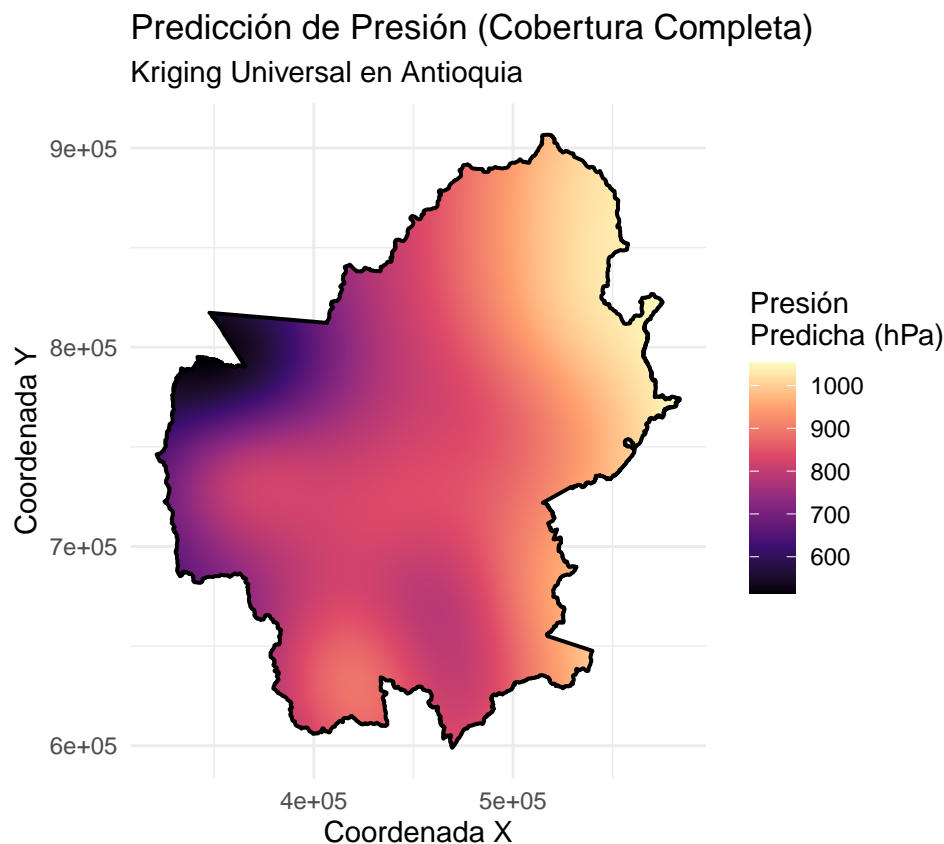
Primero, se definirá una rejilla de puntos donde se realizará la predicción y luego se aplicará el modelo de Kriging con los parámetros obtenidos.

```
## [using universal kriging]
```

```
## Warning: attribute variables are assumed to be spatially constant throughout  
## all geometries
```

### Gráfico de predicción de presión

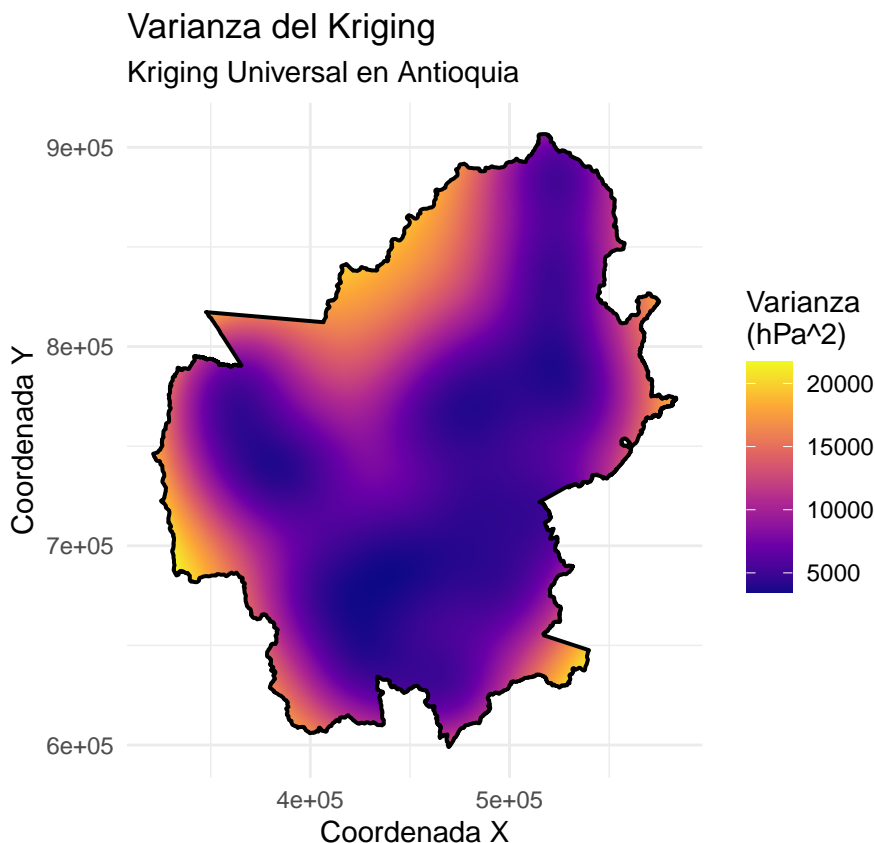
Una vez realizado el Kriging Universal, se procede a graficar la predicción de presión atmosférica en la región de Antioquia mediante un mapa de calor de las presiones predichas, superponiendo el contorno del área de estudio y los puntos originales de las estaciones.



El uso del Kriging Universal ayudó a capturar la tendencia espacial (la cual se aprecia de forma horizontal) de las presiones atmosféricas en Antioquia, proporcionando una estimación de las presiones en áreas no muestreadas. Donde se observa que las presiones más altas se encuentran en el noreste de región observada, mientras que las más bajas están en el oeste. Esto es consistente con la geografía de la región, donde las montañas y valles influyen en las presiones atmosféricas.

## Gráfico de varianza del Kriging

La varianza del Kriging proporciona una medida de la incertidumbre asociada a las predicciones. A continuación, se graficará la varianza del Kriging Universal, mostrando cómo varía la incertidumbre en diferentes áreas de Antioquia.



En el caso de la varianza, en la mayoría del mapa se observa una varianza baja, lo que indica que las predicciones son bastante confiables. Sin embargo, en algunas áreas, especialmente en los bordes del mapa, la varianza es más alta, lo que sugiere una mayor incertidumbre en las predicciones de presión atmosférica. Esto puede deberse a la menor densidad de estaciones de medición en esas áreas o a la presencia de características geográficas que afectan las presiones atmosféricas.

## 7. Validación cruzada

Para evaluar la capacidad predictiva del modelo, se aplicará una validación cruzada. En este caso, se utilizará el método Leave-One-Out Cross-Validation (LOOCV). Este proceso se repetirá para cada punto del conjunto de datos.

```
## Simple feature collection with 6 features and 6 fields
## Geometry type: POINT
## Dimension: XY
## Bounding box: xmin: 367019.2 ymin: 680633.6 xmax: 518056.4 ymax: 749369.4
## Projected CRS: WGS 84 / UTM zone 18N
##   var1.pred var1.var observed residual    zscore fold
## 1  824.7325  5835.029  792.1250 -32.60753 -0.42687073    1
## 2  659.9471  5762.708  867.3600  207.41295  2.73226336    2
```

```
## 3  938.1642 10278.874 919.8200 -18.34420 -0.18093640    3
## 4  838.1978  4929.966 809.4400 -28.75776 -0.40957462    4
## 5  846.5378  8738.359 889.8000  43.26224  0.46280076    5
## 6  796.7600  4254.791 792.2246  -4.53534 -0.06952976    6
##
##          geometry
## 1 POINT (391961.4 740631.2)
## 2 POINT (367019.2 749369.4)
## 3 POINT (518056.4 716664.3)
## 4 POINT (471955.4 697694.6)
## 5 POINT (495694.4 680633.6)
## 6 POINT (452856.8 681862.1)
```

Donde se compararán los valores observados con los valores predichos por el modelo mediante 3 métricas de error: MSE (Error Cuadrático Medio), RMSE (Raíz del Error Cuadrático Medio) y MAPE (Error Porcentual Absoluto Medio).

```
## MSE: 10187.5
```

```
## RMSE: 100.9331
```

```
## MAPE: 8.188379 %
```

- El RMSE indica que el error típico del modelo es de 100.9 hPa, lo que significa que, en promedio, las predicciones del modelo se desvían de los valores observados en aproximadamente 100.9 hPa. Considerando la escala de la presión atmosférica, este valor sugiere que el modelo tiene una precisión razonable en sus predicciones.
- El MAPE (un poco más intuitivo), indica que el modelo tiene un error porcentual medio de 8.18% en sus predicciones, lo que sugiere una buena precisión en la estimación de las presiones atmosféricas.

## Conclusiones

En este estudio, se ha desarrollado un modelo geoestadístico para predecir la presión atmosférica en regiones de Antioquia donde se tenía acceso a datos de estaciones de medición. A través del análisis estructural de distancias, se identificó una dependencia espacial significativa entre las estaciones. Este análisis también evidenció ciertas tendencias tanto en las coordenadas x e y, las cuales no se pudieron controlar con la diferenciación debido a que rompían la estructura espacial hallada, finalmente esto corrobora la necesidad de usar el Kriging Universal para la predicción.

## Apendice

```
library(tidyverse)
library(readr)
library(sf)
library(sp)
library(gstat)
library(vegan)
library(geoR)
ggplot2::theme_set(theme_minimal())
```

```
# datos
presion <- read_csv("presion2.csv")
```

```
presion2 <- presion %>%
  mutate(
    FechaObs = mdy_hms(
      FechaObservacion,
      tz = "America/Bogota"
    ),
    Hora = hour(FechaObs)
  ) %>%
  filter(Hora >= 12, Hora <= 16)

table(presion2$Hora)
```

```
# Promediar presion por estación
promedios_est <- presion2 %>%
  group_by(
    CodigoEstacion,
    NombreEstacion,
    Municipio,
    ZonaHidrografica,
    Latitud,
    Longitud
  ) %>%
  summarise(
    valor_promedio = mean(ValorObservado, na.rm = TRUE),
    n_obs          = n(),
    .groups        = "drop"
  )
```

```
pts_utm <- promedios_est %>%
  st_as_sf(coords = c("Longitud", "Latitud"), crs = 4326) %>%
  st_transform(32618)
```

```
subregiones <- st_read("Subregiones.geojson", quiet=TRUE)

antioquia2 <- subregiones %>%
  filter(COD_DEPTO == "05",
    !NOM_SUBREGION %in% c("URABÁ", "MAGDALENA MEDIO")) %>%
  st_transform(4326)

antioquia_outline <- antioquia2 %>%
  st_union() %>%
  st_sf(geometry = .)

pts_ll <- st_transform(pts_utm, crs = 4326)
```

```
ggplot() +
  geom_sf(
    data = antioquia_outline,
```



```

    fill = "gray95",
    color = "#444444",
    linewidth = 0.8
) +

geom_sf(
  data = pts_utm,
  aes(fill = valor_promedio),
  shape = 21,
  size = 4,
  color = "black",
  stroke = 0.5,
  alpha = 0.8
) +

scale_fill_distiller(
  palette = "YlOrRd",
  name = "Presión (hPa)",
  direction = 1
) +

labs(
  title = "Presión Atmosférica Promedio por Estación",
  subtitle = "Antioquia (Zonas seleccionadas)"
) +
labs(
  title = "Presión Atmosférica Promedio por Estación",
  subtitle = "Antioquia (Zonas seleccionadas)",
  x = "Este (metros)",
  y = "Norte (metros)"
) +
theme_minimal() +
coord_sf(datum = st_crs(pts_utm))

```

```

# Matriz de distancias geográficas (en metros)
dist_geo <- dist(st_coordinates(pts_utm))

# Histograma de distancias
hist(dist_geo,
     breaks = 50,
     main = "Distribución de distancias entre estaciones",
     xlab = "Distancia (m)",
     col = "skyblue")

```

```

sf_utm <- pts_utm

coords_mat <- st_coordinates(sf_utm)

dist_geo <- dist(coords_mat)

dist_var_sq <- dist(sf_utm$valor_promedio)^2

df_plot <- data.frame(

```

```

Distancia = as.vector(dist_geo),
Disimilitud = as.vector(dist_var_sq)
)

ggplot(df_plot, aes(x = Distancia, y = Disimilitud)) +
  theme_test(base_size = 14) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, color = "steelblue") +
  labs(
    x = "Distancia geográfica (m)",
    y = "Disimilitud (delta^2 de valor_promedio)",
    title = "Disimilitud vs Distancia para valor_promedio"
  ) +
  theme(
    axis.title = element_text(face = "bold"),
    axis.text = element_text(colour = "black")
  )

```

```

shp_path <- "Antioquia/Subregiones_-_Provincias_de_Colombia.shp"
municipios <- st_read(shp_path, quiet = TRUE)

```

```

crs_pts <- sf::st_crs(pts_utm)

municipios_utm <- sf::st_transform(municipios, crs_pts)

limite_utm <- sf::st_union(municipios_utm)
contorno_sf <- sf::st_boundary(limite_utm)

cont_mat <- sf::st_coordinates(contorno_sf)[,1:2]

```

```

border_df <- as.data.frame(cont_mat)
colnames(border_df) <- c("x", "y")

if (any(border_df[1, ] != border_df[nrow(border_df), ])) {
  border_df <- rbind(border_df, border_df[1, ])
}

cont_mat_sf <- st_sf(
  geometry = st_sfc(
    st_polygon(list(as.matrix(border_df)))
  )
)

# Se le asigna el mismo sistema de referencia espacial (CRS) que a los puntos
st_crs(cont_mat_sf) <- st_crs(pts_utm)

pts_utm_cuantiles <- pts_utm

pts_utm_cuantiles <- pts_utm_cuantiles %>%
  mutate(
    valor_cuantil = as.factor(ntile(valor_promedio, 4))
  )

```

```

coords <- sf::st_coordinates(pts_utm)
vals   <- pts_utm[["valor_promedio"]]

df <- data.frame(x = coords[,1],
                 y = coords[,2],
                 z = vals)

p1 <- ggplot(data = pts_utm_cuantiles) +
  geom_sf(aes(color = valor_cuantil), size = 2.5) +
  geom_sf(data = cont_mat_sf, fill = NA, color = "black", linewidth = 0.5, inherit.aes = FALSE) +
  scale_x_continuous(labels = scales::label_number(scale_cut = scales::cut_short_scale())) +
  scale_color_viridis_d(option = "plasma", name = "Cuartil") +
  coord_sf(datum = st_crs(pts_utm)) +
  labs(x = "X Coord", y = "Y Coord", color = "Valor Promedio") +
  theme_bw()

p2 <- ggplot(df, aes(x = z, y = y)) +
  geom_point(alpha = 0.8) +
  geom_smooth(method = "loess", se = FALSE, color = "black") +
  labs(x = "data", y = "Y Coord") +
  theme_bw()

p3 <- ggplot(df, aes(x = x, y = z)) +
  geom_point(alpha = 0.8) +
  geom_smooth(method = "loess", se = FALSE, color = "black") +
  labs(x = "X Coord", y = "data") +
  theme_bw()

p4 <- ggplot(df, aes(x = z)) +
  geom_histogram(aes(y = after_stat(density)),
                 color = "white", fill = "grey", bins = 8) +
  geom_density(color = "black", linewidth = 1) +
  geom_rug() +
  labs(x = "data", y = "Density") +
  theme_bw()

(p1 | p2) / (p3 | p4)

```

```

spdf <- as(pts_utm, "Spatial")
z    <- spdf@data[["valor_promedio"]]

shapiro_raw <- shapiro.test(z)
print(shapiro_raw)

qqnorm(z,
       main = "QQ-plot de valor_promedio (datos crudos)")
qqline(z, col = "red", lwd = 2)

```

```

g <- gstat(formula = valor_promedio ~ 1, data = pts_utm)

vario_emp <- variogram(g,
                      cutoff = 20000000,
                      width  = 7000,

```

```

        cressie = TRUE)

plot(vario_emp,
     main = "Semi-variograma empírico de presiones promedio",
     xlab = "Distancia (m)",
     ylab = "Semivarianza")

vgm_h <- variogram(valor_promedio ~ 1, data = pts_utm)
plot(vgm_h, main = "Semi-variograma de presión promedio")

# 1. Semi-variograma empírico
vario_emp_robust <- variogram(
  valor_promedio ~ 1,
  data      = pts_utm,
  cutoff    = 49000,
  width     = 2000,
  cressie   = F
)

# 2. Simulaciones permutadas con estimador
nsim <- 99
vg_sim_robust <- map_dfr(seq_len(nsim), ~{
  tmp <- pts_utm %>% mutate(z = sample(valor_promedio))
  variogram(
    z ~ 1,
    data  = tmp,
    cutoff = 49000,
    width  = 2000,
    cressie = F
  ) %>%
  mutate(sim = .x)
})

# 3. Construcción de la envolvente (2.5% y 97.5%)
env_df_robust <- vg_sim_robust %>%
  group_by(dist) %>%
  summarise(
    lower = quantile(gamma, .025),
    upper = quantile(gamma, .975)
  )

# 4. Preparar datos para el plot uniendo empírico y envolvente
vg_plot_robust <- vario_emp_robust %>%
  left_join(env_df_robust, by = "dist")

# 5. Dibujar con ggplot2
ggplot(vg_plot_robust, aes(x = dist)) +
  geom_ribbon(aes(ymin = lower, ymax = upper),
            fill = "grey80", alpha = .6) +
  geom_line(aes(y = gamma), colour = "blue", size = 1) +
  geom_point(aes(y = gamma), colour = "darkblue", size = 1.5) +
  labs(
    x      = "Distancia (m)",

```

```

    y      = "Semivarianza",
    title = "Semi-variograma con envolvente "
  ) +
  theme_minimal()

set.seed(987)

dist_att <- dist(promedios_est$valor_promedio)
mantel_res <- mantel(
  dist_geo,
  dist_att,
  permutations = 999,
  method = "pearson"
)

mantel_res

ini.vals <- expand.grid(seq(2, 6, l=5), seq(5, 10, l=5))

coords <- st_coordinates(pts_utm)

valor <- pts_utm$valor_promedio

data_geoR <- as.geodata(cbind(coords, valor), coords.col = 1:2, data.col = 3)

# Maxima verosimilitud restringida (REML)
fit_reml <- likfit(
  geodata = data_geoR,
  cov.model = "gaussian",
  ini.cov.pars = c(56132, 216888),
  nugget = 0,
  lik.method = "REML"
)

# Maxima verosimilitud (ML)
fit_ml <- likfit(
  geodata = data_geoR,
  cov.model = "gaussian",
  ini.cov.pars = c(56132, 216888),
  nugget = 0,
  lik.method = "ML"
)

print("Resultados REML:")
print(fit_reml)

print("Resultados ML:")
print(fit_ml)

v_empirico <- variog(data_geoR, option = "bin", uvec = seq(0, 49000, by = 2000))

plot(v_empirico, main = "Ajustes con ML y REML",

```

```

    xlab = "Distancia (m)", ylab = "Semivarianza")

lines.variomodel(fit_reml, col = "blue", lwd = 2)

lines.variomodel(fit_ml, col = "red", lwd = 2)

pts_utm <- pts_utm %>%
  mutate(
    x = st_coordinates(.)[,1],
    y = st_coordinates(.)[,2]
  )

rejilla_grande <- st_make_grid(limite_utm, cellsize = 1000, what = "centers")

rejilla_grande_sf <- st_as_sf(rejilla_grande) %>%
  rename(geometry = x) %>%
  mutate(
    x = st_coordinates(.)[,1],
    y = st_coordinates(.)[,2]
  )

modelo_final_uk <- vgm(psil = 10590, model = "Gau", range = 48000, nugget = 2886.7)

kriging_completo <- krige(
  formula = valor_promedio ~ x + y,
  locations = pts_utm,
  newdata = rejilla_grande_sf,
  model = modelo_final_uk
)

kriging_recortado_final <- st_intersection(kriging_completo, limite_utm)

df_grid <- kriging_recortado_final %>%
  mutate(
    x = st_coordinates(.)[, 1],
    y = st_coordinates(.)[, 2]
  )

df_grid <- kriging_recortado_final %>%
  mutate(x = st_coordinates(.)[,1],
         y = st_coordinates(.)[,2]) %>%
  as.data.frame()

contorno_df <- contorno_sf %>%
  st_coordinates() %>%
  as.data.frame()

pts_df <- as.data.frame(pts_utm)

ggplot() +
  geom_raster(data = df_grid, aes(x = x, y = y, fill = var1.pred)) +
  geom_path(data = contorno_df, aes(x = X, y = Y), color = "black", linewidth = 0.7) +

```

```

scale_fill_viridis_c(option = "magma", name = "Presión\nPredicha (hPa)") +
labs(
  title = "Predicción de Presión",
  subtitle = "Kriging Universal en Antioquia",
  x = "Coordenada X",
  y = "Coordenada Y"
) +
coord_equal() +
theme_minimal()

```

```

ggplot() +
  geom_raster(data = df_grid, aes(x = x, y = y, fill = var1.var)) +
  geom_path(data = contorno_df, aes(x = X, y = Y), color = "black", linewidth = 0.7) +
  scale_fill_viridis_c(option = "plasma", name = "Varianza\n(hPa^2)") +
  labs(
    title = "Varianza del Kriging",
    subtitle = "Kriging Universal en Antioquia",
    x = "Coordenada X",
    y = "Coordenada Y"
  ) +
  coord_equal() +
  theme_minimal()

```

```

validacion_cruzada <- krige.cv(
  formula = valor_promedio ~ x + y,
  locations = pts_utm,
  model = modelo_final_uk,
  nfold = nrow(pts_utm) # Para hacer LOOCV
)

head(validacion_cruzada)

```

```

mse <- mean(validacion_cruzada$residual^2)
rmse <- sqrt(mse)
mape <- mean(abs(validacion_cruzada$residual / validacion_cruzada$observed)) * 100
cat("MSE:", mse, "\n")
cat("RMSE:", rmse, "\n")
cat("MAPE:", mape, "%\n")

```