UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Facultad de Estadística
Maestría en Estadística Aplicada
Junio 2025

ANÁLISIS DE LA CORRELACIÓN ESPACIAL DE LAS PRECIPITACIONES EN BOGOTÁ, MEDIANTE LA PRUEBA I DE MORAN PARA DATOS FUNCIONALES

ANALYSIS OF THE SPATIAL CORRELATION OF RAINFALL IN BOGOTÁ, USING MORAN'S I TEST FOR FUNCTIONAL DATA

Pedro Nel Miranda Rivera.[a]
pedromiranda@usantotomas.edu.co

Wilmer Pineda Rios.[b]
wilmerpineda@usantomas.edu.co

Resumen

Este trabajo de grado es un proyecto cuantitativo y descriptivo, que se basa en los datos de las precipitaciones registradas en las 91 estaciones de pluviometría que el IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales), tiene en la ciudad de Bogotá, con datos recogidos entre los años de 2018 a 2022. Al Observar el conjunto de datos disponibles de las 91 estaciones, se hace necesario reducir el número de estaciones a 19, debido a que tienen la información más consistente tanto, por el mayor número de registros realizados, como el menor número de NA ("Not Available"(No Disponible)). Al medir y analizar la correlación espacial existente entre dichas estaciones, mediante una modificación de la prueba I de Moran definida para datos espaciales y aplicarla a los datos funcionales que resultan de esta base depurada, se concluye que no existe correlación espacial entre las estaciones, cuando se estudia la variable precipitación. El anterior resultado se comprobó utilizando 5 formas diferentes de medir las distancias entre las estaciones, en el conjunto de las 19 estaciones en general y en los 5 subconjuntos en particular que se estudiaron.

***Palabras clave*:** Prueba I de Moran, correlación espacial, datos funcionales, pluviometría, IDEAM.

Abstract

This degree work is a quantitative and descriptive project, based on data on precipitation data recorded at the 91 pluviometric stations operated by IDEAM (Institute of Hydrology, Meteorology and Environmental Studies) in the city of Bogotá, with data collected between 2018 and 2022. Upon reviewing the dataset from the 91 stations, it became necessary to reduce the number of stations to 19, as these provided the most consistent data—both in terms of the higher number of recorded observations and the lower incidence of NA ("Not Available") values. By measuring and analyzing the spatial correlation among these selected stations using a modified version of Moran's I test—adapted for spatial data and applied to the functional data derived from this refined dataset—it was concluded that there is no spatial correlation between the stations when analyzing the precipitation variable. This result was confirmed using five different methods of measuring distances between stations, both across the overall set of 19 stations and within five specific subsets that were examined.

***Keywords*:** Moran's I test, spatial correlation, functional data, pluviometry, IDEAM.

[a] Estudiante
[b] Director

1. Introduccíon

En este trabajo se mide la correlación espacial entre 19 estaciones seleccionadas de pluviometría del IDEAM, en la ciudad de Bogotá, teniendo como variable de estudio, la precipitación medida en mm y con los registros recolectados entre los años 2018 y 2022, esto mediante de una modificación de la prueba I de Moran diseñada para datos espaciales y aplicada en el contexto de datos funcionales. Se define como unidad de análisis la ciudad de Bogotá y como unidades de observación las 19 estaciones que poseen los datos más completos. Para lograr lo anterior, en primer lugar, se estudia la prueba I de Moran para datos espaciales en el capítulo 4, a continuación, se expone la teoría de datos funcionales en general en el capítulo 5 y luego en el capítulo 6 la prueba I de Moran para datos funcionales en particular. En seguida en el capítulo 7 se trabaja el tratamiento del conjunto de datos, luego, en el capítulo 8 se realiza el cálculo e interpretación de la prueba I de Moran, tanto para los datos discretos iniciales, como para las curvas funcionales finales, se aplica la prueba I de Moran a la base depurada e imputada para medir la correlación espacial y para confirmar los resultados obtenidos, se cambia de método de medición de las distancias entre las estaciones y además se estudia el comportamiento de la correlación espacial en diferentes subgrupos de las estaciones según su cercanía. Finalmente, en el capítulo 9 se exponen las conclusiones y recomendaciones, al final del documento se encuentran las referencias y los anexos correspondientes.

Como este trabajo se basa en los registros de las precipitaciones en la ciudad de Bogotá, es conveniente aclara unas definiciones básicas:

1.1. PLUVIOMETRÍA O PLUVIMETRÍA

Continuando con Miller (2019)la pluviometria es la parte de la meteorología que se encarga de medir la cantidad, intensidad y regularidad de las lluvias o precipitaciones en una zona geográfica y tiempo determinados. Su principal herramienta de medición es el pluviómetro, instrumento que recoge y mide las precipitaciones en un lugar y tiempos determinados, generalmente se encuentra instalado en las estaciones meteorológicas. Las lluvias o precipitaciones se miden en milímetros (mm) y corresponde al espesor de una lámina de agua sobre una base de un metro cuadrado plana e impermeable, equivalente a un litro respecto a un metro cuadrado (1mm = 1L / m2).

Precipitación: Se considera como cualquier tipo de condensación del vapor de agua atmosférico que se deposita sobre la superficie terrestre y se consideran tres tipos de precipitaciones:

- Precipitación líquida: llovizna y lluvia
- Precipitación glacial: llovizna congelada y lluvia congelada (aguanieve)
- Precipitación congelada: nieve, bolitas de nieve, granos de nieve, bolitas de hielo, granizo, bolitas o copos de nieve y cristales de hielo
- Precipitación líquida: llovizna y lluvia

Figura 1: Recuperado de: https://www.instagram.com/p/CeynUTUu93O/

- Precipitación glacial: llovizna congelada y lluvia congelada (aguanieve)

Figura 2: Recuperado de: https://www.minutouno.com/sociedad/nieve/ola-frio-que-diferencia-hay-agua-graupel-y-lluvia-helada-n6027543

- Precipitación congelada: nieve, bolitas de nieve, granos de nieve, bolitas de hielo, granizo, bolitas o copos de nieve y cristales de hielo

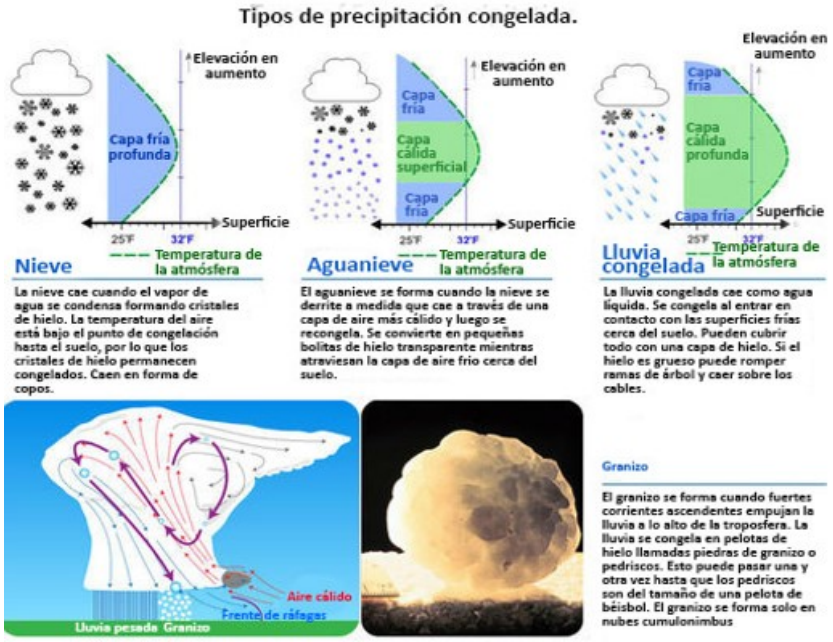


Figura 3: Recuperado de: https://flexbooks.ck12.org/cbook/ck-12-conceptos-de-ciencias-de-la-tierra-grados-6-8-en-espanol/section/8.5/primary/lesson/precipitaciones/

Características de la precipitación

Tamaño y forma:
Regularmente las gotas de lluvia tienen diámetros que oscilan entre 0.1 mm a 9 mm Las más pequeñas se denominan gotitas de nube y son esféricas.

Intensidad y duración:
Regularmente la relación entre intensidad y duración de las precipitaciones es inversa, es decir, las lluvias de intensidad altas probablemente serán de duración corta y las lluvias de intensidad baja pueden tener una duración larga.

Intensidad y área:
Generalmente sobre un área grande las precipitaciones suelen ser menos intensa que sobre un área pequeña.

Tamaño de gota e intensidad
Las lluvias de intensidad alta tienen un tamaño de gota más grande que las tormentas de intensidad baja.

Y de acuerdo con Ward, R. C., Robinson, M. (2000). Los datos sobre precipitaciones son más fáciles de obtener, para más lugares y durante períodos más largos, que para otros componentes del ciclo del agua. Es por esta razón que es de suma importancia realizar trabajos que cuantifiquen el comportamiento de las precipitaciones en las diferentes ciudades y así poder sumar las conclusiones que puedan dar cuenta de ésta fenómeno vital para el ser humano, en medio del cambiante panorama del cambio climático.

Además, como la fuente de los datos es el IDEAM, se expone un resumen de la institución:

1.2. IDEAM

En concordancia con lo recuperado de `http://www.ideam.gov.co/` 2024, IDEAM es la sigla que corresponde al Instituto de Hidrología, Meteorología y Estudios Ambientales y es una entidad del gobierno de

Colombia dependiente del Ministerio de Ambiente y desarrollo sostenible. Su objetivo es el manejo de la información científica respecto a la climatología, meteorología, hidrología y todo lo relacionado con el medio ambiente en Colombia. Su fecha de creación fue el 22 de diciembre de 1993, cuando el Congreso de la República sancionó la ley 99 de 1993, con la cual se creó el Ministerio del Medio Ambiente y además se reemplazó al Instituto Colombiano de Hidrología, Meteorología y Adecuación de Tierras (HIMAT), por el IDEAM iniciando su funcionamiento oficialmente el 1 de marzo de 1995.

También tiene la función de administrar el funcionamiento y ubicación de las bases meteorológicas e hidrológicas dentro del país, con el fin de recopilar información, pronósticos, alertas y asesoría sobre el comportamiento del clima a la población.

El Instituto tiene como misión generar conocimiento y garantizar el acceso a la información sobre el estado de los recursos naturales y condiciones hidrometeorológicas de todo el país para la toma de decisiones de la población, autoridades, sectores económicos y sociales de Colombia.

El IDEAM tiene una red de 91 estaciones en la ciudad de Bogotá, en el anexo 1 se presenta la relación de sus nombres, ubicación y altitud.

Por otro lado, dentro de la literatura encontrada respecto a estudios similares donde se utiliza el índice I de Moran, se pueden observar los siguientes estudios:

- Desarrollo de un Sistema de Información Geográfico (SIG) para el análisis de patrones espaciales de incendios en viviendas. Celestino Ordoñez Galán, María Rosa Varela González, Aimara Reyes Pantoja Departamento de Ingeniería de los Recursos Naturales y Medio Ambiente (IRNMA), Escuela Universitaria de Ingeniería Técnica Industrial, Rúa Torrecedeira, 86, 36208 Vigo (España). 2011
- Autocorrelación espacial: analogías y diferencias entre el índice de moran y el índice getis y ord. aplicaciones con indicadores de acceso al agua en el norte argentino. Dra. Liliana Ramírez Mgtr. Vilma Lilián Falcón Departamento e Instituto de Geografía. Facultad de Humanidades. Universidad Nacional del Nordeste Consejo Nacional de Investigaciones Científicas y Técnicas – CONICET.2008
- Econometría espacial y el análisis sociodemográfico. Aplicación en la formación de agrupaciones espaciales de envejecimiento en Cuba, período 2003-2009. Dra. Otilia Barros Díaz Ph. Patricio Aroca González. 2009

En todos los trabajos anteriores se utiliza la prueba I de Morán para la medida de la correlación espacial para datos netamente espaciales, en este trabajo se utiliza su modificación para aplicarla a datos funcionales.

Respecto a la literatura encontrada con estudios similares donde se buscan objetivos parecidos, se pueden observar los siguientes trabajos:

- Análisis de la distribución e interpolación espacial de las lluvias en Bogotá, Colombia. Andrés Vargas Departamento de Ingeniería Civil, Pontificia Universidad Javeriana. Ana santos universidad nacional de Colombia. Pontificia universidad javeriana. Eder cárdenas Estudiante Maestría en Hidrosistemas, Pontificia Universidad Javeriana. Nelson obregón Pontificia Universidad Javeriana. 2011
- Estudio de la caracterización climática de Bogotá y cuenca alta del río Tunjuelo. IDEAM y FOPAE 2007

En los documentos anteriores se realizan estudios sobre las lluvias realizando interpolaciones espaciales, o análisis de las precipitaciones es una determinada zona de la ciudad, a diferencia con este trabajo que estudia la corrección espacial de las estaciones con mayores registros entre los años 2018 y 2022.

En cuanto a trabajos realizados aplicando la técnica de análisis de datos funcionales, se puede apreciar:

- Análisis de Datos Funcionales aplicado a datos de temperatura en España, David Miguel Picón Llamas, Facultad de Ciencias, Departamento de Estadística, Universidad de Valladolid, 2019.

2. OBJETIVOS

2.1. OBJETIVO GENERAL

Medir la correlación espacial de las precipitaciones en la ciudad de Bogotá, con los registros recolectados entre los años 2018 y 2022, en las 19 estaciones seleccionadas de pluviometría del IDEAM, mediante la modificación de la prueba I de Moran que permite aplicarla a datos funcionales específicamente y no a datos espaciales.

2.2. OBJETIVOS ESPECIFICOS

1. Analizar la modificación de la prueba I de Moran definida respecto a datos espaciales, para aplicarla en datos funcionales.
2. Depurar e Imputar la base de datos de precipitaciones disponible en el IDEAM, que es de dominio público y consta de 12 variables o columnas y aproximadamente 194.000.000 de registros o filas.
3. Utilizar la prueba I de Moran de correlación espacial para el contexto de datos funcionales, en la base depurada.
4. Realizar el cambio de método de medición de distancias, requerido en la prueba I de Moran para datos funcionales, de acuerdo a 5 modelos y comprobar la similitud de los resultados.

3. PLANTEAMIENTO DEL PROBLEMA

¿Existe correlación espacial entre las 19 estaciones pluviométricas seleccionadas de la ciudad de Bogotá, de acuerdo a los registros de lluvias comprendidos entre los años 2018 a 2022, de acuerdo a la prueba I de Moran para datos funcionales?

Como lo mencionan Vargas A. y otros en su artículo ANÁLISIS DE LA DISTRIBUCIÓN E INTERPOLACIÓN ESPACIAL DE LAS LLUVIAS EN BOGOTÁ, COLOMBIA Universidad Javeriana 2.011, es de suma importancia el conocimiento de la distribución espacial de las precipitaciones en la ciudad de Bogotá para diseñar el sistema de drenaje urbano, en ese sentido este trabajo tiene un valor preponderante.

4. INTRODUCCIÓN A LA ESTADÍSTICA ESPACIAL

La estadística espacial estudia variables que están indexadas por ubicaciones geográficas. A diferencia de otros dominios, la proximidad espacial entre observaciones introduce dependencia que debe ser modelada explícitamente. Existen tres grandes clases de datos espaciales:

- **Geoestadísticos**: Datos en ubicaciones continuas y fijas (p.ej., concentración de minerales, temperatura). Se modelan a través de funciones de covarianza y semivariogramas.
- **Datos de Área o Lattice Data**: Datos agregados en unidades espaciales discretas (p.ej., municipios, celdas de imagen). Se utilizan modelos autorregresivos espaciales.
- **Procesos Puntuales**: Ubicaciones aleatorias de eventos (p.ej., incendios, crímenes, rayos). Se analizan con métodos de funciones de intensidad y estadísticos espaciales como la función K de Ripley.

4.1. Aplicaciones comunes

- **Meteorología**: La geoestadística se ha empleado extensamente en la modelación de fenómenos atmosféricos como la temperatura, precipitación, velocidad del viento, humedad del suelo y detección de rayos (Berrocal (2010), Cameletti(2013)).
- **Epidemiología**: Identificación de focos de enfermedades y su relación con condiciones ambientales (Lawson (2013))
- **Ciencias ambientales**: Predicción de contaminación del aire, calidad del agua y dispersión de contaminantes (Jerrett (2005))
- **Economía regional**: Modelado de desigualdad socioeconómica, pobreza y desempleo (Anselin1998).
- **Criminología**: Detección de zonas calientes y análisis de patrones delictivos.
- **Agronomía y recursos naturales**: Estimación de rendimiento de cultivos y características del suelo.

4.2. Representación de la Matriz de Proximidad

En estadística espacial, la matriz de pesos espaciales
$mathbfW = [w_{ij}]$ cumple un papel central en la modelación de la dependencia entre unidades geográficas. Su función principal es cuantificar la estructura de vecindad o similitud espacial entre observaciones, permitiendo incorporar explícitamente la ubicación en el análisis estadístico.

Objetivo de la Matriz de Pesos

El objetivo de construir esta matriz es representar formalmente la idea de proximidad o relación entre unidades espaciales, de modo que se puedan identificar patrones de autocorrelación (positiva o negativa), realizar interpolaciones, o ajustar modelos espaciales como SAR, CAR o kriging. La elección adecuada de la matriz de pesos afecta de forma crítica tanto la detección de dependencia espacial como la robustez de los modelos subsecuentes.

4.3. Criterios de Construcción

La matriz puede definirse de diferentes maneras, dependiendo del tipo de datos y de la escala geográfica:

1. Datos de Área (Lattice Data)

En estos casos, las unidades espaciales están definidas sobre una estructura discreta (p.ej., municipios, celdas), y la vecindad se determina mediante relaciones topológicas:

- **Contigüidad tipo torre (rook)**: Dos regiones son vecinas si comparten una frontera (borde).
- **Contigüidad tipo reina (queen)**: Dos regiones son vecinas si comparten frontera o vértice.

Ventajas:

- No requieren coordenadas geográficas, solo la estructura del mapa.
- Capturan adecuadamente la relación administrativa o política entre regiones.
- Fáciles de implementar con polígonos vectoriales (shapefiles).

2. Datos Geoestadísticos

Cuando se dispone de coordenadas exactas en un espacio continuo, es común definir los pesos mediante funciones de distancia entre ubicaciones s_i y s_j:

$$d_{ij} = |s_i - s_j| \tag{1}$$

Y se transforma en pesos w_{ij} usando funciones de decaimiento espacial:

- **Distancia inversa**: $w_{ij} = 1/d_{ij}^k$, con $k > 0$.
- **Kernel Gaussiano**: $w_{ij} = \exp(-d_{ij}^2/(2h^2))$
- **Kernel Exponencial**: $w_{ij} = \exp(-d_{ij}/h)$
- **Kernel Epanechnikov**: $w_{ij} = \frac{3}{4}(1 - (d_{ij}/h)^2)$ si d_{ij} leh, y 0 si no.

Ventajas:

- Incorporan suavizamiento espacial flexible con un parámetro de ancho de banda h.
- Permiten modelar relaciones de proximidad en entornos irregulares o no contiguos.
- Aptos para interpolaciones (kriging) y para procesar grandes bases de datos con sensores distribuidos.

Características Adicionales de W

- Puede ser **simétrica** ($w_{ij} = w_{ji}$) o no, dependiendo del criterio.
- Puede estar **estandarizada por filas** (cada fila suma 1), útil para modelos de tipo SAR.
- Es común que sea **dispersa** (la mayoría de $w_{ij} = 0$), lo que permite almacenamiento eficiente y reducción de complejidad computacional.

Consideraciones Prácticas

La selección del esquema de pesos espaciales no debe ser arbitraria; debe reflejar tanto la teoría del fenómeno bajo estudio como la resolución espacial disponible. Por ejemplo:

- En epidemiología o economía regional, las relaciones de contigüidad suelen ser más significativas.
- En meteorología o análisis ambiental, las funciones continuas con kernels permiten capturar mejor la gradualidad y dispersión de los fenómenos.

La matriz **W** actúa como un componente estructurante en la mayoría de los modelos espaciales, y su especificación adecuada mejora sustancialmente la capacidad explicativa y predictiva del análisis espacial.

4.4. Índice de Moran

El Índice de Moran (I) es una medida clásica de autocorrelación espacial global, análoga al coeficiente de correlación de Pearson, pero adaptada para capturar patrones espaciales en los datos. Se define como:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(Z_i - \bar{Z})(Z_j - \bar{Z})}{\sum_{i=1}^{n}(Z_i - \bar{Z})^2} \tag{2}$$

Donde:

- n es el número total de unidades espaciales (observaciones).
- Z_i es el valor observado de la variable de interés en la unidad i.
- $\bar{Z}$ es la media global de la variable: $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$.
- w_{ij} es el peso espacial entre las unidades i y j, normalmente definido a partir de una matriz de contigüidad o función de distancia.

El numerador del índice evalúa la covarianza espacial ponderada entre los valores Z_i y Z_j, mientras que el denominador estandariza la medida dividiendo por la varianza total. El índice I puede interpretarse de la siguiente manera:

- $I > 0$: indicativo de autocorrelación espacial positiva (valores similares tienden a agruparse).
- $I < 0$: indicativo de autocorrelación espacial negativa (valores disímiles se encuentran próximos).
- $I \approx 0$: sugiere ausencia de autocorrelación espacial (distribución aleatoria).

Juzgamiento de hipótesis

Para evaluar la significancia estadística del índice de Moran, se formula el siguiente contraste de hipótesis:

$$H_0: \quad \text{No existe autocorrelación espacial} \quad (I \approx \mathbb{E}[I])$$

$$H_1: \quad \text{Existe autocorrelación espacial} \quad (I \neq \mathbb{E}[I])$$

Bajo la hipótesis nula de aleatoriedad espacial (sin dependencia), se puede calcular el valor esperado del índice y su varianza (asumiendo normalidad e independencia):

$$\mathbb{E}[I] = -\frac{1}{n-1} \tag{3}$$

A partir de esta expectativa, se construye la siguiente estadística tipo Z:

$$Z_I = \frac{I_{\text{obs}} - \mathbb{E}[I]}{\sqrt{\text{Var}(I)}} \tag{4}$$

Donde $\text{Var}(I)$ puede obtenerse de forma analítica (dependiendo del tipo de pesos y del supuesto de normalidad), o empíricamente a través de simulaciones.

Métodos de Inferencia

Existen dos enfoques principales para juzgar la hipótesis:

1. **Prueba basada en permutaciones:** se generan muchas permutaciones aleatorias de los valores observados Z_i, manteniendo fija la matriz de pesos W. Se calcula I para cada permutación, y se obtiene una distribución empírica de referencia. El valor-p se estima como la proporción de permutaciones con un índice tan extremo como el observado.
2. **Prueba normal aproximada:** se calcula la estadística Z_I y se compara contra la distribución normal estándar. Se rechaza H_0 si $|Z_I| > z_{\alpha/2}$ para un nivel de significancia α (por ejemplo, $z_{0.025} = 1.96$ para $\alpha = 0.05$).

Este contraste permite evaluar si existe una estructura espacial no aleatoria en los datos, lo cual es fundamental para aplicar métodos espaciales adecuados y evitar errores en la inferencia convencional.

4.5. Coeficiente de Geary

El Coeficiente de Geary (C) es otra medida de autocorrelación espacial, pero a diferencia del Índice de Moran, es más sensible a variaciones locales. Mientras que Moran evalúa la covariación global, Geary se enfoca en las diferencias absolutas entre valores vecinos. Se define como:

$$C = \frac{(n-1)}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(Z_i - Z_j)^2}{\sum_{i=1}^{n}(Z_i - \bar{Z})^2} \tag{5}$$

Donde:

- n es el número de unidades espaciales.
- Z_i y Z_j son los valores observados en las ubicaciones i y j respectivamente.
- $\bar{Z}$ es la media global de la variable.
- w_{ij} representa el peso espacial entre las ubicaciones i y j.

La estructura de esta fórmula destaca las diferencias cuadráticas entre observaciones vecinas, ponderadas por la matriz de pesos. A diferencia del Índice de Moran, que tiene valores teóricos en el intervalo $[-1, 1]$, el Coeficiente de Geary toma valores en el rango $[0, 2]$, y se interpreta así:

- $C < 1$: indica autocorrelación espacial positiva (valores similares están cerca).
- $C > 1$: indica autocorrelación espacial negativa (valores disímiles están cerca).
- $C \approx 1$: sugiere ausencia de autocorrelación espacial (aleatoriedad).

Hipótesis de Prueba

Para determinar si el valor observado de C indica dependencia espacial significativa, se plantea el siguiente contraste:

$$H_0: \quad \text{No existe autocorrelación espacial} \quad (C \approx \mathbb{E}[C] = 1)$$

$$H_1: \quad \text{Existe autocorrelación espacial} \quad (C \neq 1)$$

Métodos de Inferencia

Existen dos métodos principales para evaluar la significancia de C:

1. **Permutaciones aleatorias:** Se reordenan aleatoriamente los valores de Z manteniendo fija la estructura de la matriz de pesos W, y se calcula C para cada permutación. La distribución empírica obtenida permite calcular un valor-p comparando C_{obs} con esta distribución nula.
2. **Aproximación normal:** Bajo supuestos de normalidad e independencia, se puede aproximar la distribución de C mediante una distribución normal, utilizando su esperanza y varianza teóricas. Se define una estadística tipo Z como:

$$Z_C = \frac{C_{obs} - \mathbb{E}[C]}{\sqrt{\text{Var}(C)}}, \quad \mathbb{E}[C] = 1 \tag{6}$$

   El valor crítico se compara con los percentiles de la distribución normal estándar. Se rechaza H_0 si $|Z_C| > z_{\alpha/2}$.

En comparación con el Índice de Moran, el Coeficiente de Geary puede detectar variaciones locales sutiles, lo cual es especialmente útil cuando existen "bordes" o transiciones abruptas entre regiones con diferentes valores. Por esta razón, ambos indicadores se consideran complementarios en el análisis exploratorio espacial.

4.6. Fundamentos de la Geoestadística

La geoestadística es una rama de la estadística espacial orientada al análisis y modelado de fenómenos que varían de manera continua en el espacio geográfico. A diferencia de los datos de área o de procesos puntuales, los datos geoestadísticos surgen de observaciones realizadas en ubicaciones geográficas específicas, dentro de un dominio continuo $D \subset \mathbb{R}^2$ o $\mathbb{R}^3$.

El objetivo de la geoestadística es predecir o interpolar el valor de una variable aleatoria espacialmente correlacionada en ubicaciones no muestreadas, utilizando la información de puntos vecinos. Esta predicción se fundamenta en modelar la estructura de dependencia espacial entre observaciones mediante funciones de covarianza o semivariogramas.

Procesos Espaciales

Un proceso espacial puede definirse como una colección de variables aleatorias $\{Z(s), s \in D\}$, donde s denota una localización espacial. Cada $Z(s)$ representa el valor de la variable aleatoria en la ubicación s.

Estacionariedad y Tipos de Dependencia

La inferencia geoestadística se basa en ciertas suposiciones sobre la estructura del proceso:

- **Estacionariedad de primer orden:** Se cumple si la media del proceso es constante en todo el dominio:
$$\mathbb{E}[Z(s)] = \mu \quad \forall s \in D$$
- **Estacionariedad de segundo orden (o débil):** Se cumple si la covarianza entre dos ubicaciones sólo depende de su separación $h = s_i - s_j$ y no de su ubicación absoluta:
$$\text{Cov}(Z(s), Z(s+h)) = C(h)$$
- **Isotropía:** Se cumple si la covarianza depende únicamente de la distancia euclidiana $||h||$ entre puntos, es decir, la dirección no afecta la correlación:
$$C(h) = C(||h||)$$
- **Anisotropía:** Cuando la dependencia espacial varía con la dirección, por ejemplo, si el viento o la pendiente afectan más en un eje que en otro.

Función de Covarianza

La covarianza espacial $C(h)$ describe la correlación entre valores de la variable Z separados por un vector h:

$$C(h) = \text{Cov}(Z(s), Z(s+h))$$

Es importante que $C(h)$ sea una función positiva definida para garantizar la validez del modelo estadístico.

4.7. Función de Semivariograma

El **semivariograma** mide la disimilitud media esperada entre pares de observaciones separadas por una distancia h:

$$\gamma(h) = \frac{1}{2}\mathbb{E}[(Z(s) - Z(s+h))^2]$$

La relación entre la función de covarianza y el semivariograma es:

$$\gamma(h) = C(0) - C(h)$$

Estimación Empírica del Semivariograma

Dado que el verdadero semivariograma es desconocido, se estima a partir de los datos mediante el semivariograma empírico:

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(i,j)\in N(h)} (Z(s_i) - Z(s_j))^2$$

donde $N(h)$ es el conjunto de pares de observaciones cuya separación se encuentra en una ventana alrededor de h.

Componentes del Semivariograma Teórico

El semivariograma teórico suele representarse mediante modelos paramétricos como el esférico, exponencial o gaussiano. Los siguientes parámetros son fundamentales:

- **Nugget** (τ^2): representa la variabilidad a escalas espaciales muy pequeñas (incluyendo error de medición o micro-variación no captada). Se interpreta como el salto en el semivariograma en $h \approx 0$.
- **Sill** ($\sigma^2 + \tau^2$): es el valor al que se estabiliza el semivariograma, correspondiente a la varianza total del proceso.
- **Range** (r): es la distancia a partir de la cual ya no hay correlación espacial; es decir, el semivariograma alcanza el sill.
- **Modelos típicos**:
  - **Exponencial:** $\gamma(h) = \tau^2 + \sigma^2(1 - \exp(-h/\phi))$
  - **Gaussiano:** $\gamma(h) = \tau^2 + \sigma^2(1 - \exp(-(h/\phi)^2))$
  - **Esférico:**

$$\gamma(h) = \begin{cases} \tau^2 + \sigma^2 \left[\frac{3h}{2r} - \frac{1}{2}\left(\frac{h}{r}\right)^3\right], & h \leq r \\ \tau^2 + \sigma^2, & h > r \end{cases}$$

Interpolación Espacial: Kriging

Una de las principales aplicaciones de la geoestadística es el **kriging**, un método de predicción espacial óptimo (en el sentido de minimizar el error cuadrático medio) que utiliza el semivariograma ajustado. La predicción de $Z(s_0)$ en una ubicación no observada s_0 se realiza como combinación lineal de las observaciones:

$$\hat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i)$$

Los pesos λ_i se determinan resolviendo un sistema lineal que incorpora la estructura de covarianza o semivariograma, asegurando que la predicción sea insesgada y de varianza mínima.

De acuerdo con Kopczewska (2020), El índice I de Morán es una medida global de la autocorrelación espacial, es decir mide si existe cercanía o no entre los valores de una variable determinada para sitios ubicados en una región específica, lo que permite identificar conglomerados y relaciones espaciales en la

región. Su valor va de -1 a 1 ($-1 \leq I \leq 1$), en forma estricta, para valores de Z-score menores -1,96 o superiores a 1,96 establecen autocorrelación espacial con un nivel de significancia de 0,05

Teniendo en cuenta a Xavier (2010), se establece que si al estudiar un conjunto de datos de una variable específica, recogido en diferentes lugares, se le asocia las coordenadas geográficas correspondientes, su resultado será mucho más completo que si no se tuviera en cuenta su ubicación.

Según el espacio euclidiano tradicional de tres dimensiones (x, y, z) en el cual se está midiendo una variable "a" se puede trabajar de dos maneras diferentes: "a" se puede trabajar de dos maneras diferentes:

- Estudiar sólo la distribución espacial independientemente de "a", es decir cómo se disponen las ternas (x,y,z)
- Analizar toda la información, tanto la disposición espacial y los valores recogidos, conjuntos de (x,y,z,a)

Los procedimientos básicos del análisis estadístico espacial son los siguientes:

- Medidas centro-gráficas. Son aquellas similares a las medidas de tendencia central como la (media) o la mediana, y también las de dispersión como la desviación típica.
- Análisis estadístico de líneas. Descriptores estadísticos para líneas y ángulos.
- Análisis de patrones de puntos. Estudia la estructura espacial de un conjunto de puntos en función de parámetros como la densidad o las distancias entre puntos y su configuración en el espacio
- Autocorrelación espacial. Establece la veracidad del postulado que afirma que los puntos cercanos tienden a tener valores más similares entre sí que los puntos alejados.

INTERPRETACIÓN

Siendo el índice "I" de Moran un indicador de estadística deductiva, sus resultados se interpretan respecto a la hipótesis nula, que establece que el atributo estudiado se distribuye en forma aleatoria respecto a las entidades de estudio.

Siendo el valor P estadísticamente significativo (5 %), se interpreta con siguientes resultados:

Valor P	Interpretación
No es estadísticamente significativo	No se rechaza la hipótesis nula. Es posible que la distribución espacial de los valores de entidades sea el resultado de procesos espaciales aleatorios.
Es estadísticamente significativo y la puntuación z es positiva.	Puede rechazar la hipótesis nula. La distribución espacial de los valores altos y los valores bajos en el conjunto de datos está más agrupada espacialmente de lo que se esperaría si los procesos espaciales subyacentes fueran aleatorios
Es estadísticamente significativo y la puntuación z es negativa.	Puede rechazar la hipótesis nula. La distribución espacial de los valores altos y los valores bajos en el conjunto de datos está más dispersa espacialmente de lo que se esperaría si los procesos espaciales subyacentes fueran aleatorios.

Tabla 1: Interpretaciones del valor P para la prueba I de Moran Recuperado de: `https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatialstatistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm`

De acuerdo con Martori, J.C. (1999) existen 4 criterios de vecindad que se deben tener en cuenta para construir la matriz W de pesos espaciales, a saber:

- Criterio lineal: Se denominan vecinas de i aquellas regiones que comparten el lado izquierdo o derecho de i
- Criterio Torre: Se denominan vecinas de i aquellas regiones que comparten algún lado de i
- Criterio Alfil: Se denominan vecinas de i aquellas regiones que comparten algún vértice de i
- Criterio Reina: Se denominan vecinas de i aquellas regiones que comparten algún lado o vértice de i

Según Anselin (1988) la matriz de pesos espaciales W, es aquella matriz cuadrada, con valores positivos finitos, con valores en diagonal igual a cero, que representa la interdependencia de los lugares geográficos asociados, los valores diferentes de cero, corresponden a unidades espaciales contiguas. Esta matriz W establece la longitud de todas las interacciones posibles entre los sitios geográficos trabajados. Puede representarse W, de la siguiente manera:

$$w = \begin{bmatrix} 0 & w_{1,2} & \dots & w_{1,N} \\ w_{2,1} & 0 & \dots & w_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N,1} & w_{N,2} & \dots & 0 \end{bmatrix} \tag{7}$$

A continuación, se presentan diferentes gráficos obtenidos a partir de la base de datos discretos, es decir antes de transformarlos a datos funcionales o curvas continuas.

En la figura 4 se observan las coordenadas geográficas de las 19 estaciones de pluviometría del IDEAM de la ciudad de Bogotá, seleccionadas para el estudio.

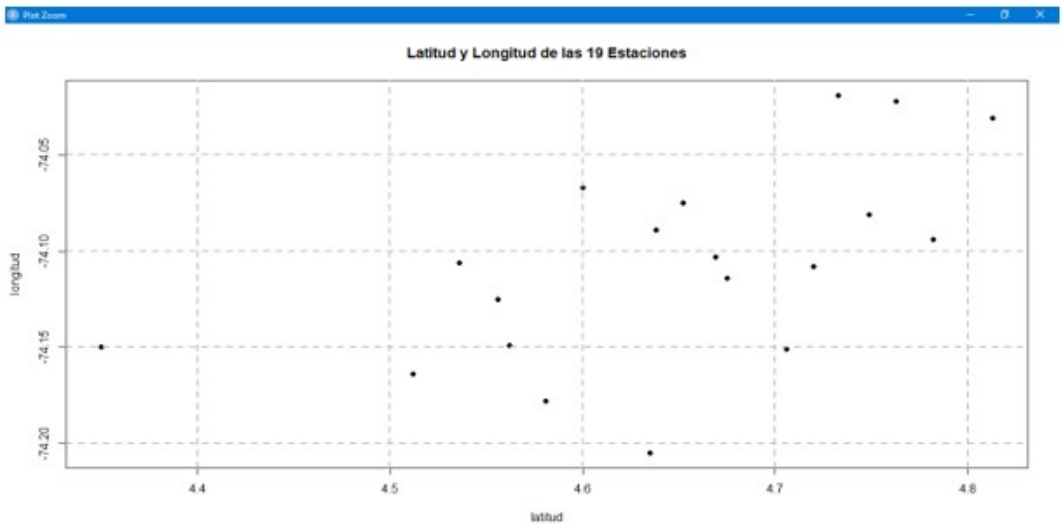


Figura 4: Gráfico de la distribución espacial de las 19 estaciones de pluviometría del IDEAM de Bogotá seleccionadas para el estudio. Elaboración propia

En la figura 5 se observan la latitud de las 19 estaciones de pluviometría del IDEAM de la ciudad de Bogotá, seleccionadas para el estudio, como variables de acuerdo al paquete geoR, utilizado para georeferencias

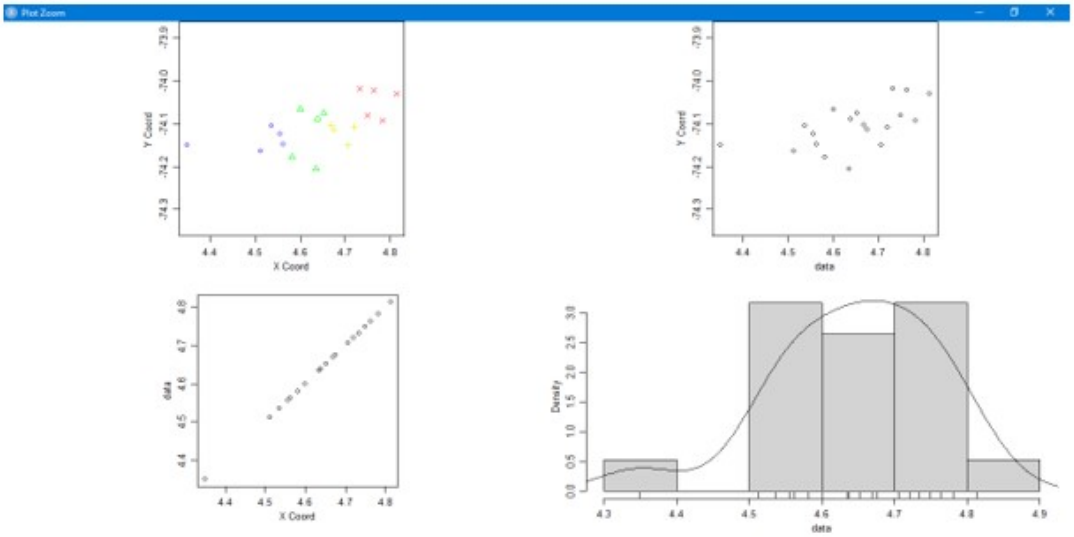


Figura 5: Gráfico de la latitud de las 19 estaciones de pluviometría del IDEAM de Bogotá como variable tipo geoR. Elaboración propia

En la figura 6 se observan la longitud de las 19 estaciones de pluviometría del IDEAM de la ciudad de Bogotá, seleccionadas para el estudio, como variables de acuerdo al paquete geoR, utilizado para georeferencias.

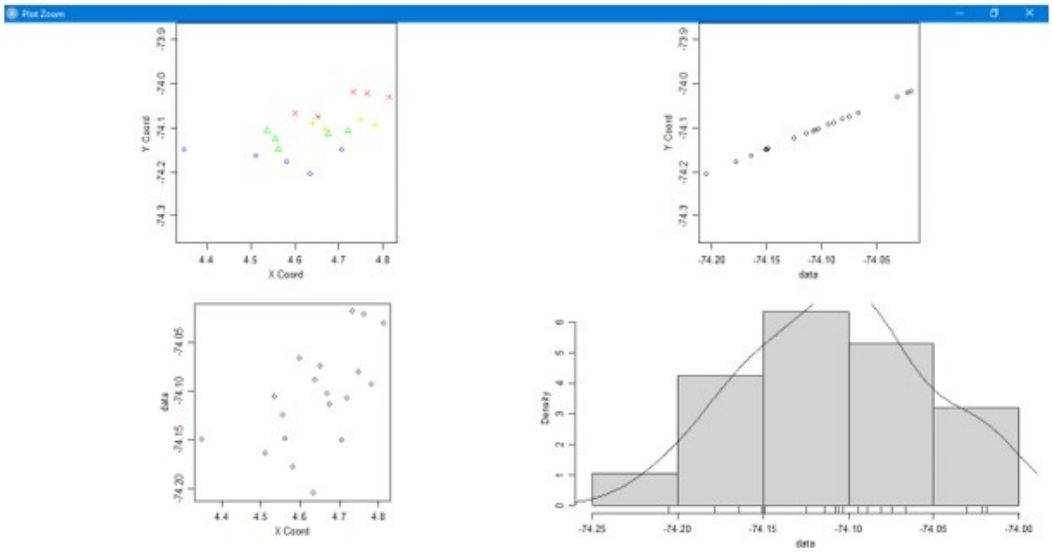


Figura 6: Gráfico de la longitud de las 19 estaciones de pluviometría del IDEAM de Bogotá como variable tipo geoR. Elaboración propia

En concordancia con Schabenberger, O. y Pierce, f. (2001), una herramienta habitual para captar la estructura de segundo momento en los datos espaciales es el semivariograma. Es un gráfico que permite analizar cómo cambia alguna característica respecto a la distancia, cuando el proceso espacial es intrínseco y, además, estacionario de segundo orden, se puede definir respecto a la función de covarianza $C(s_i - s_j)$ así:

$$\gamma(s_i - s_j) = C(0) - C(s_i - s_j) \tag{8}$$

Donde:

- $\gamma(s_i - s_j)$ es el calculo del semivariograma entre la característica "s" medida en el punto "i" y entre la característica "s" medida en el punto "j"

- $C(0)$ es la covarianza calculada en el punto 0

- $C(s_i - s_j)$ es la covarianza de la diferencia entre la característica "s" medida en el punto "i" y la característica "s" medida en el punto "j"

De acuerdo con Armstrong, A. (1998). Observando los principales elementos de un variograma, se presentan varios de estos generados a partir del conjunto de datos discretos trabajado.

En la figura 7 se puede apreciar que en el variograma de la latitud de las 19 estaciones pluviométricas seleccionadas no existe el elemento meseta, ni el rango o alcance y el umbral es ligeramente mayor a 0,10.

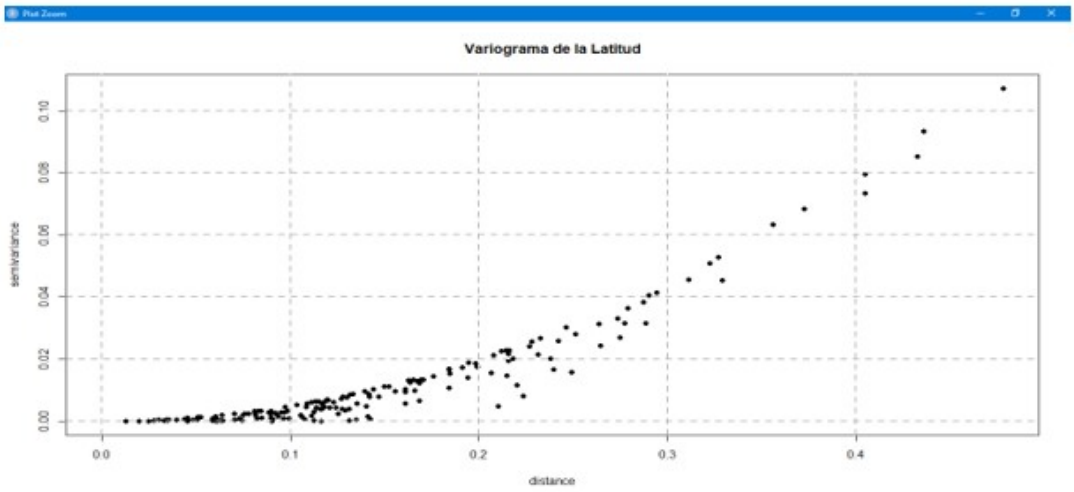


Figura 7: Variograma de la latitud de las 19 estaciones de pluviometría del IDEAM de Bogotá como variable tipo geoR. Elaboración propia

En la figura 8 se puede apreciar que en el variograma de la longitud de las 19 estaciones pluviométricas seleccionadas no existe el elemento meseta, ni el rango o alcance y el umbral es ligeramente mayor a 0,015.

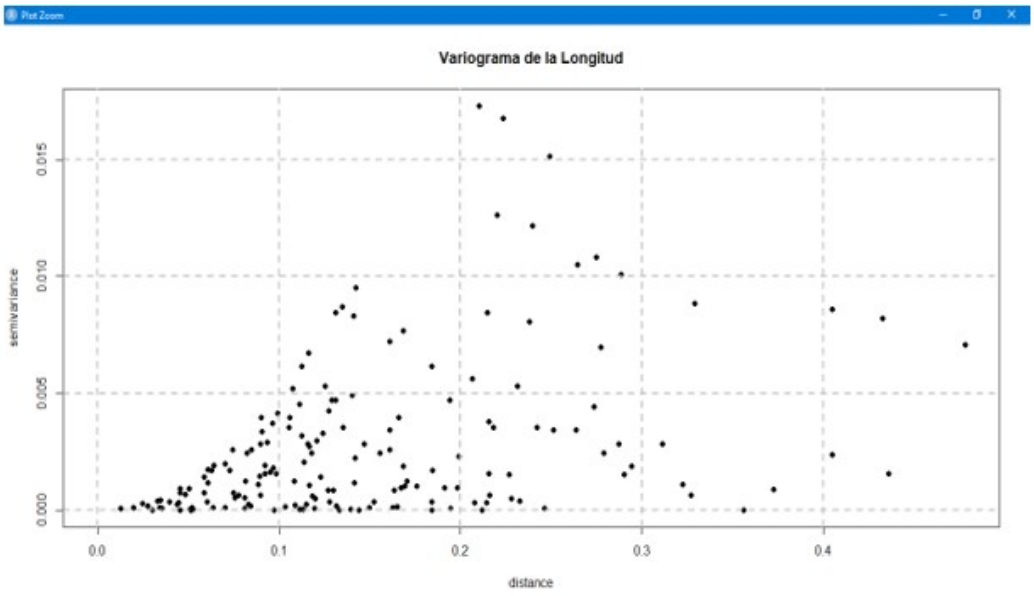


Figura 8: Variograma de la longitud de las 19 estaciones de pluviometría del IDEAM de Bogotá como variable tipo geoR. Elaboración propia

En la figura 9 se puede apreciar que, en el variograma de los datos de las 19 estaciones pluviométricas seleccionadas en el 1° día de estudio, no existe el elemento meseta, ni el rango o alcance y el umbral es ligeramente mayor a 0,005.

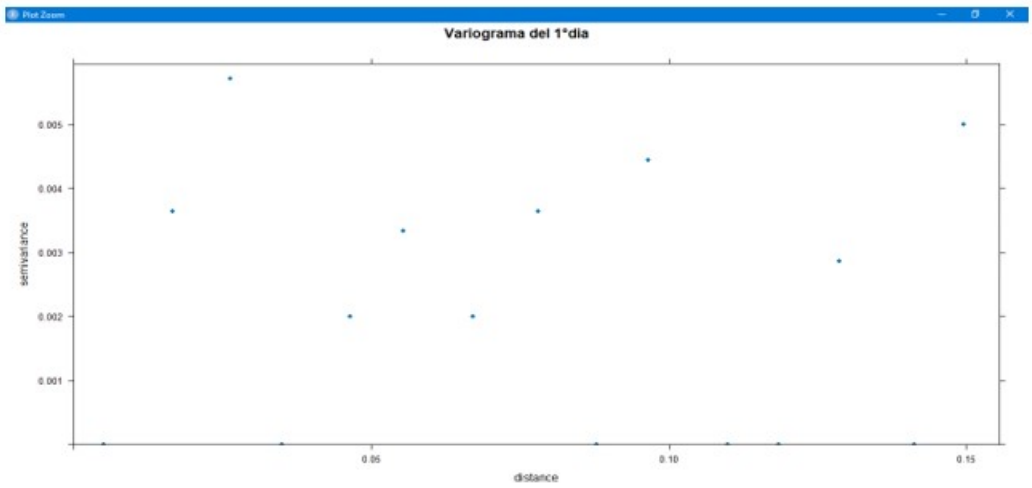


Figura 9: . Variograma de las 19 estaciones de pluviometría del IDEAM de Bogotá el 1° día de estudio. Elaboración propia

En la figura 10 se puede apreciar que, en el variograma de los datos de las 19 estaciones pluviométricas seleccionadas en el 2° día de estudio, no existe el elemento meseta, ni el rango o alcance y el umbral es ligeramente mayor a 15.

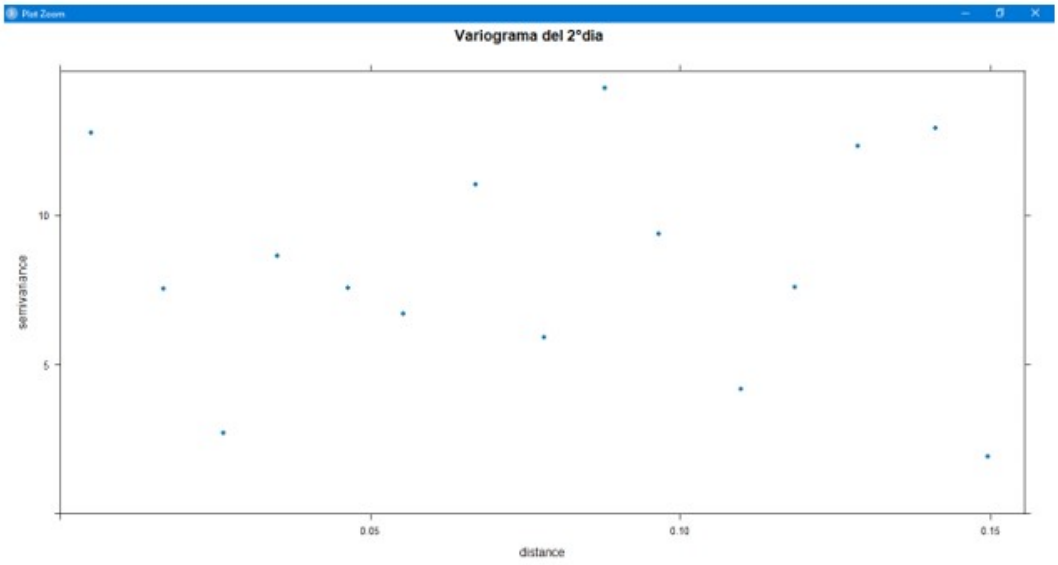


Figura 10: Variograma de las 19 estaciones de pluviometría del IDEAM de Bogotá el 2° día de estudio. Elaboración propia

En la figura 11 se puede apreciar que, en el variograma de los datos de las 19 estaciones pluviométricas seleccionadas en el 3° día de estudio, no existe el elemento meseta, ni el rango o alcance y el umbral es ligeramente mayor a 6.

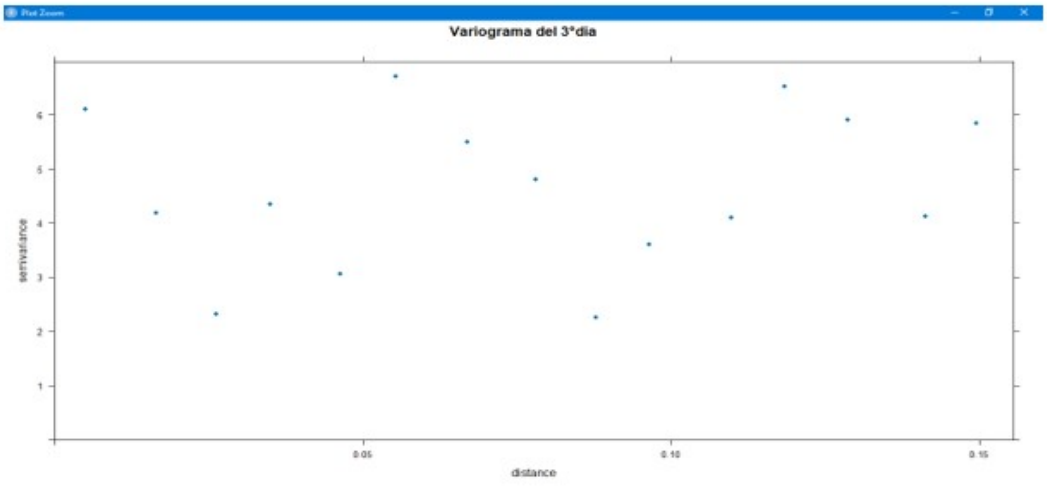


Figura 11: Variograma de las 19 estaciones de pluviometría del IDEAM de Bogotá el 3° día de estudio. Elaboración propia

En la figura 12 se puede apreciar que, en el variograma de los datos de las 19 estaciones pluviométricas seleccionadas en el 4° día de estudio, no existe el elemento meseta, ni el rango o alcance y el umbral es ligeramente mayor a 25.

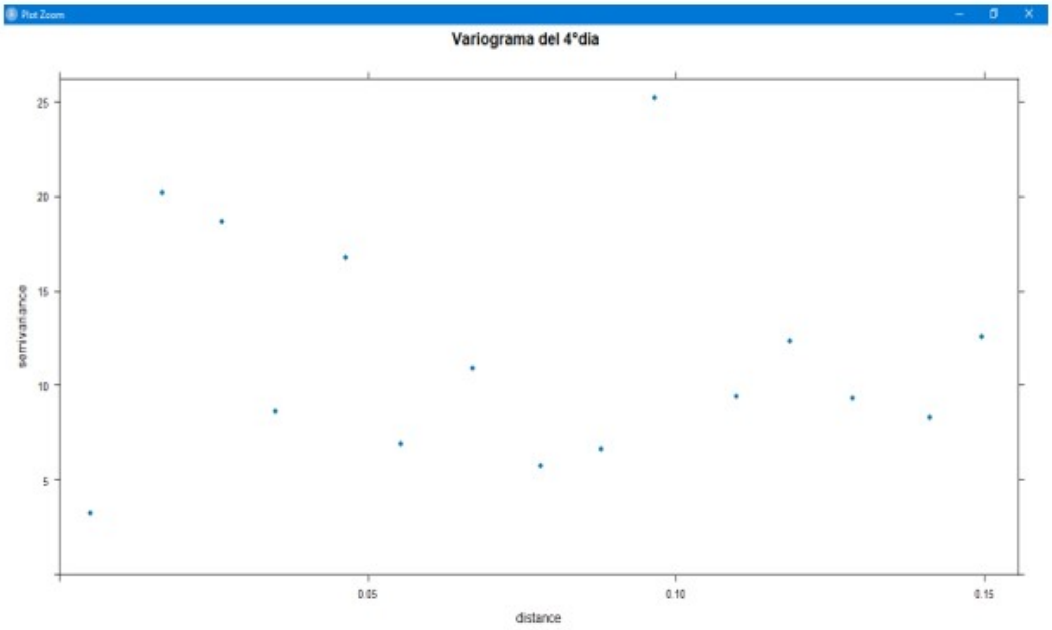


Figura 12: Variograma de las 19 estaciones de pluviometría del IDEAM de Bogotá el 4° día de estudio. Elaboración propia

Y así sucesivamente se pueden realizar variogramas en los 1232 días de la base de datos trabajados. Se puede concluir por tanto que después de observar estos variogramas, no se puede determinar en forma específica, alguna región de dependencia espacial

5. DATOS FUNCIONALES

5.1. DEFINICION:

Según Horvath L. (2012) la constante proliferación de información respecto a diferentes aspectos de la vida diaria en contextos geográficos diferentes hace que el análisis de datos funcionales "FDA" por sus siglas en inglés, sea una herramienta fundamental para el estudio y comprensión de la realidad actual. Sin embargo, quien por primera vez utilizó los términos de "Análisis de datos funcionales" fue el estadístico James O. Ramsay (1982), aunque las bases de este enfoque ya habían sido establecidas en los trabajos de Grenander (1950) y Rao (1958)

En el presente trabajo, se aplica esta técnica para establecer si existe o no correlación espacial entre los datos de las precipitaciones en la ciudad de Bogotá, se toman como referencia las 19 estaciones de pluviometría del IDEAM que mejores registros tienen. Para ello se hace una revisión básica de la teoría respectiva. El análisis de datos funcionales es una técnica utilizada para estudiar la relación que existe entre una variable independiente y una o más variables dependientes, basándose en el análisis de funciones matemáticas para modelar y entender cómo cambian los datos a medida que varía la variable independiente.

Los datos funcionales son observaciones de procesos estocásticos definidas en un intervalo ininterrumpido de tiempo, se pueden entender como curvas continuas, no como los datos discretos que son los que regularmente se trabajan

Los datos funcionales pueden ser definidos como:

$$x_i(t),\ t \in\ T = [a,b] \subset R \tag{9}$$

Se puede representar por el par: (t_j, x_{ij}) con $t_j \in\ T,\ j = 1, 2, \ldots,\ N$,

Entonces un conjunto de datos funcionales es la observación $x_1, x_2, \ldots, x_n$ de n variables funcionales distribuidas como X

donde N es la cantidad de puntos donde se observa la variable de interés $y_{i,j} = X_i(t_j) + \varepsilon_j$

Lo anterior representa una serie de datos visto en N funciones de una variable durante un periodo de tiempo continuo y espacio determinado, Donde el espacio de Hilbert $L^2(\tau)$ (Entendido como la generalización del espacio Euclidiano para dimensiones infinitas) está definido como el espacio de funciones al cuadrado integrables dadas en un proceso estocástico

De acuerdo con Raya (2007) los espacios de Hilbert son aquellos que tienen definido un producto interior $(X\langle\bullet,\bullet\rangle)$ y un espacio métrico(X,d)completo, con

$$d(x,y) = \|x-y\|(x,y \in X)$$

Y

$$\|x\| = \langle x,x\rangle^{\frac{1}{2}}$$

Y en cada caso individual, la curva se puede expresar como:

$$x_n(t) : t \in [T_1, T_2],\quad n = 1, 2, \ldots., N \tag{10}$$

Al ser $x_n(t_{j,n})$ cualquier elemento de $L^2(\tau)$ se cumple que:

$$||E|| = E\left[\int X^2(t)\,dt\right]^{\frac{1}{2}} < \infty \tag{11}$$

Por otro lado, es conveniente tener en cuenta la siguiente definición:

Las variables funcionales se caracterizan por la evolución de su valor a lo largo del tiempo (proceso estocástico), de modo que los valores que toman son funciones en lugar de vectores como en análisis multivalente clásico. Existe una problemática manifiesta cuando se quiere medir en forma continua este tipo variables, además se observa una complejidad teórica de los métodos estadísticos para analizarlas, por esto se generan resúmenes periódicos que constituyen las series temporales. La mayoría de los modelos predictivos los datos temporales discretos, requieren fuertes restricciones, observaciones igualmente espaciadas o pertenencia a una clase de procesos específica.

De lo anterior se puede determinar como variables funcionales las curvas formadas por los valores dados en mm, de las precipitaciones registradas en cualquiera de las 19 estaciones durante los 1232 días registrados en la base de datos del IDEAM y se manifiesta la necesidad de pasar esos datos discretos a curvas continuas para realizar un análisis más general.

5.2. EXPANCIONES EN BASES:

Lo primero que se debe hacer en cada curva es expresarla por medio de una función base:

$$X_n(t) \approx \sum_{m=1}^{M} c_{nm} B_m(t), \quad 1 \leq n \leq N \tag{12}$$

Donde:

$X_n(t)$ son las curvas que se están trabajando

c_{nm} son los coeficientes asociados para realizar la combinación lineal

Y los $B_m(t)$ son diferentes conjuntos preestablecidos de funciones como seno, coseno o *B-Splines* o Fourier. En la siguiente figura se pueden apreciar algunos ejemplos:

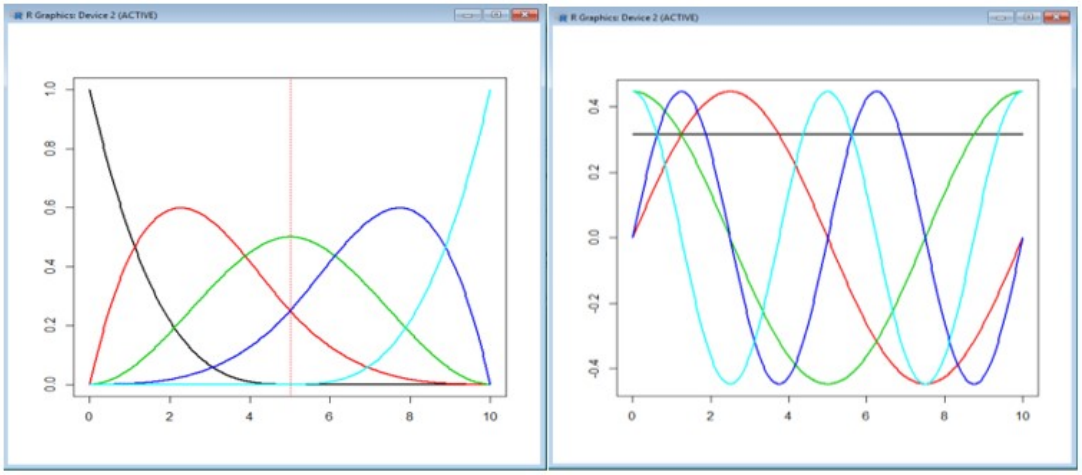


Figura 13: Gráficos de las primeras cinco funciones básicas *B-splines* (izquierda) y Fourier (derecha). Adaptación de Kokoszka, P. Reimherr, M .2017

Según Ramsay and Silverman (2005) "las funciones básicas son como ondas o splines, o funciones seno y coseno que comparten algunas propiedades de forma, por lo que pueden aproximarse como combinaciones lineales de algunas M formas básicas."

5.3. SUAVIZAMIENTO

Es la técnica mediante la cual se puede pasar de datos discretos a curvas y que requieren que cumplan los siguientes requerimientos:

- Que se pueda realizar el registro en una escala de tiempo común
- Poder obtener la lectura del valor de la variable en cualquier momento del periodo de tiempo establecido
- Valorar las tasas de cambio en cualquier momento
- Reducir el ruido

De acuerdo con el profesor Rubén D. Guevara G. (2015): En la vida real las observaciones funcionales son observaciones discretas en n puntos. Entonces, la primera tarea es convertir estos valores discretos en una función. Si los valores discretos son asumidos sin error, entonces el proceso se llama interpolación, pero si ellos tienen algún error observacional, este debe ser considerado y el proceso se denomina suavizamiento.

Para el suavizamiento se tienen dos opciones: si los datos son periódicos se utilizan combinaciones lineales de bases de Fourier, de lo contrario como en este estudio, se trabajan combinaciones lineales de Bases B-Spline, pues al realizar las gráficas de las precipitaciones en las 19 estaciones, no se observan variaciones cíclicas fuertes. Para lograr funciones suaves es decir que posean una o más derivadas. En este estudio se trabaja el sistema de Bases B-Spline creado por De Boor, C. (2001) que es al más conocido y está disponible en la mayoría de lenguajes de programación estadísticos, incluyendo R.

Se requiere de una base de funciones linealmente independientes denominadas como ϕ_k que permitan ajustar las observaciones discretas asi: Sean $y_j, \quad j = 1, \ldots$,n los datos discretos

$y_j = x(t_j) + \varepsilon_j$ con

$$x(t) = \sum_{k}^{K} C_k \phi_k(t) = c'\phi \tag{13}$$

Donde:

$x(t)$ es una curva de las observadas

$\phi_k(t)$ son las funciones linealmente independientes

C_k representan los coeficientes que garanticen la independencia lineal

Para esto se realiza el siguiente procedimiento:

- El primer paso para definir un *spline* es dividir el intervalo sobre el cual la función está definida en L subintervalos. Los puntos de separación son llamados *knots*.
- Sobre cada intervalo, un *spline* es un polinomio de orden especificado m.
- Polinomios adyacentes son unidos suavemente en los *knots*. En estos puntos los polinomios toman el mismo valor funcional.

Teniendo en cuenta que:

- El orden del polinomio = grado del polinomio o su potencia más alta +1
- El número de grados de libertas en el ajuste = orden del polinomio + el número de *knots* interiores

- El número de funciones base = Grado del polinomio + Número de *knots*-1

Con los siguientes objetivos:

- Asegurar que la estimación de la curva de un buen resultado de los datos en términos de que la suma al cuadrado de los errores $y_j = x(t_j) + \varepsilon_j$ sea mínima
- El ajuste no puede ser exacto porque entonces la curva x es excesivamente rugosa o localmente variable

Y realizando los siguientes cálculos:

$$SMSSE\ (y|\ c) = \sum_{j=1}^{n}\left[y_j - \sum_{k}^{K} C_k\phi_k(t_j)\right]^2 = (y-\phi_c)\prime(y-\phi_c) \tag{14}$$

Derivando e igualando a cero:

$$2\phi\phi' c - 2\phi' y = 0 \tag{15}$$

$$\hat{c} = (\phi\prime\phi)^{-1}\phi' y \tag{16}$$

$$\hat{y} = \phi\hat{c} = \phi(\phi\prime\phi)^{-1}\phi' y \tag{17}$$

5.4. CAMINATA ALEATORIA:

La caminata aleatoria o paseo aleatorio, es un proceso discreto tanto en el espacio como en el tiempo y consiste en una secuencia de pasos aleatorios de tamaño específicos, pero si los pasos se hacen cada vez más pequeños, es decir su longitud tiende a 0 y el tiempo toma valores discretos muy cercanos entre sí, el límite corresponde al proceso Wiener, o movimiento browniano, que es estocástico y continuo tanto en el espacio como en el tiempo.

Para entender mejor la construcción de una base de expansión, se expondrá el ejemplo de la expansión B-spline del proceso Wiener, o movimiento Browniano, mediante el paquete Fda. en R que fue originalmente diseñado para acompañar el libro de Ramsay y Silverman (2005). Definido como:

Una función aleatoria $\{W(t), t \in [0,1]\}$ se denomina proceso de Wiener si se cumplen las siguientes condiciones:

1. $W(0) = 0$
2. Si $0 \leq s < t \leq 1$, entonces $W(t) - W(s)$ es normal con media cero y varianza $t - s$
3. Para cualquier $0 \leq t_0 < t_1 < \ldots t_k \leq 1$ las variables aleatorias $W(t_j) - W(t_{j-1}), 1 \leq j \leq k$ son independientes
4. si $\{W(t), t \in [0,1]\}$ es un proceso Wiener, entonces $B(t) = W(t) - tW(1), t \in [0,1]$ es llamado un puente Browniano.

La forma habitual de aproximar un proceso de Wiener se basa en la observación de que para Si $0 \leq k \leq K$

$$W\left(\frac{k}{K}\right) - W\left(\frac{k-1}{K}\right)\ N\left(0, \frac{1}{K}\right)\ \frac{1}{\sqrt{K}}N_k \tag{18}$$

Entonces, con la normal estandarizada independiente N_k

$$W\left(\frac{k}{K}\right) = \frac{1}{\sqrt{K}}\sum_{i=1}^{k} N_k \tag{19}$$

$$S_i = \frac{1}{\sqrt{K}} \sum_{k=1}^{i} N_k \ , \quad N_k \sim \ iid \ \ N(0,1) \, , \qquad 1 \leq k \leq K \tag{20}$$

Según Kokoszka, Reimherr (2017) La siguiente figura muestra una trayectoria del paseo aleatorio que puede ser visto como una función definida en el intervalo [0, K] para $X(t_i) = S_i, \ t_i = i$

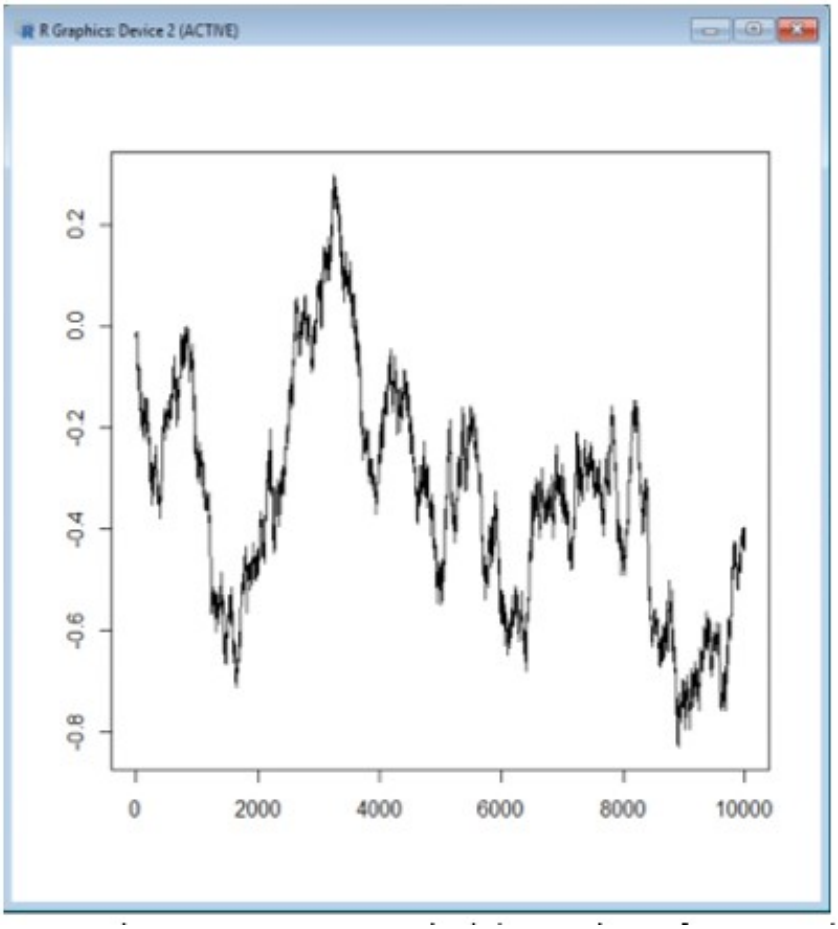


Figura 14: Caminata aleatoria y su expansión (3) usando 25 funciones bases B-spline Adaptación de Kokoszka, P. Reimherr, M .2017

5.5. MEDIA, DESVIACION ESTÁNDAR Y COVARIANZA

Ya teniendo las curvas suavizadas se procede a graficar la media y la desviación estándar mediante la aplicación de las siguientes definiciones:

Media:

$$\bar{X}_N(t) = \frac{1}{N} \sum_{n=1}^{N} X_n(t) \tag{21}$$

Donde:

$\bar{X}_N(t)$ Es la curva de la media de las N curvas

$\sum_{n=1}^{N} X_n(t)$ Es la sumatoria de los valores de las N curvas

N es el número de curvas

Desviacion estandar:

$$SD_X(t) = \left\{ \frac{1}{N-1} \sum_{n=1}^{N} \left[X_n(t) - \bar{X}_N(t) \right]^2 \right\}^{\frac{1}{2}} \tag{22}$$

Con:

$SD_X(t)$ Es la curva de la desviación estándar de las N curvas

$\left[X_n(t) - \bar{X}_N(t) \right]^2$ Es el cuadrado de diferencia de cada curva respecto a la media de todas

N es el número de curvas

Si se simulan una muestra de N = 50 paseos aleatorios y se convierten en objetos funcionales, al graficarlos, junto con su media y DE, se pueden observar en la siguiente Figura.

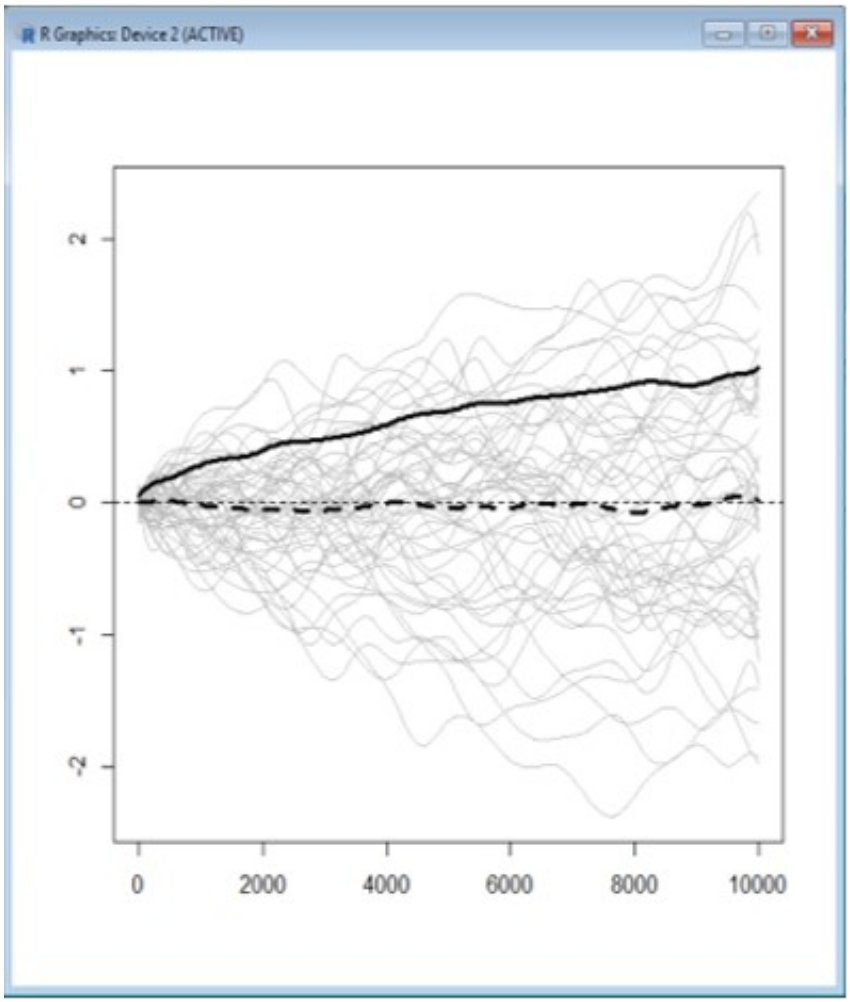


Figura 15: Paseos aleatorios convertidos en objetos funcionales junto con la desviación estándar SD (línea continua gruesa) y la media (gruesa Línea discontinua). Adaptación de Kokoszka, P. Reimherr, M .2017

Covarianza:

$$\widehat{c}(t,s) = \frac{1}{N-1} \sum_{n=1}^{N} \{(X_n\}(t) - \bar{X}_N(t))(X_n(s) - \bar{X}_N(s)) \tag{23}$$

La interpretación de los valores de $\hat{c}\,(t, s)$ es la misma que para la matriz de covarianza habitual, Por ejemplo, los valores grandes indican que Xn (t) y Xn (s) tienden a estar simultáneamente por encima o por debajo de los valores promedio en estos puntos. Si se consideran los 50 paseos aleatorios convertidos en objetos funcionales anteriormente se pueden generar las siguientes gráficas de perspectiva y contorno de la covarianza.

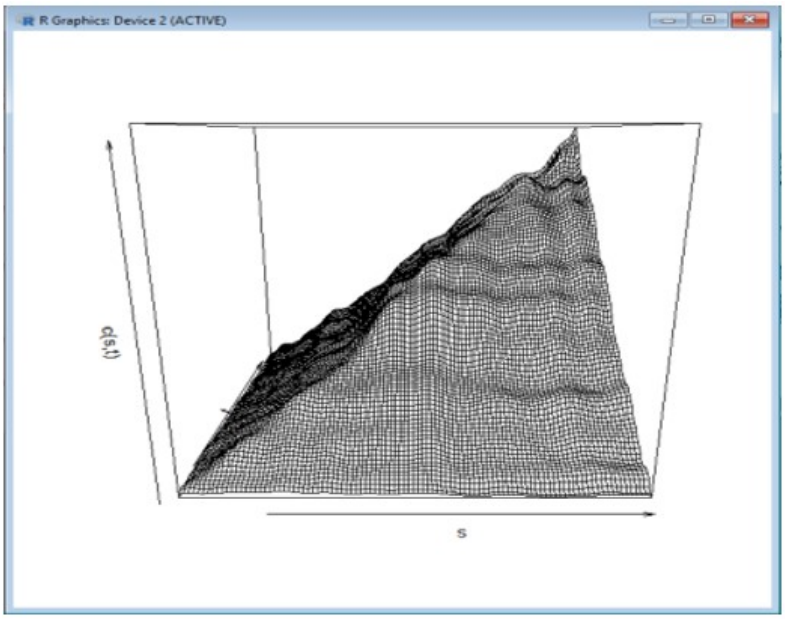


Figura 16: Un diagrama de perspectiva de la covarianza función de la muestra de 50 caminatas aleatorias Adaptación de Kokoszka, P. Reimherr, M .2017

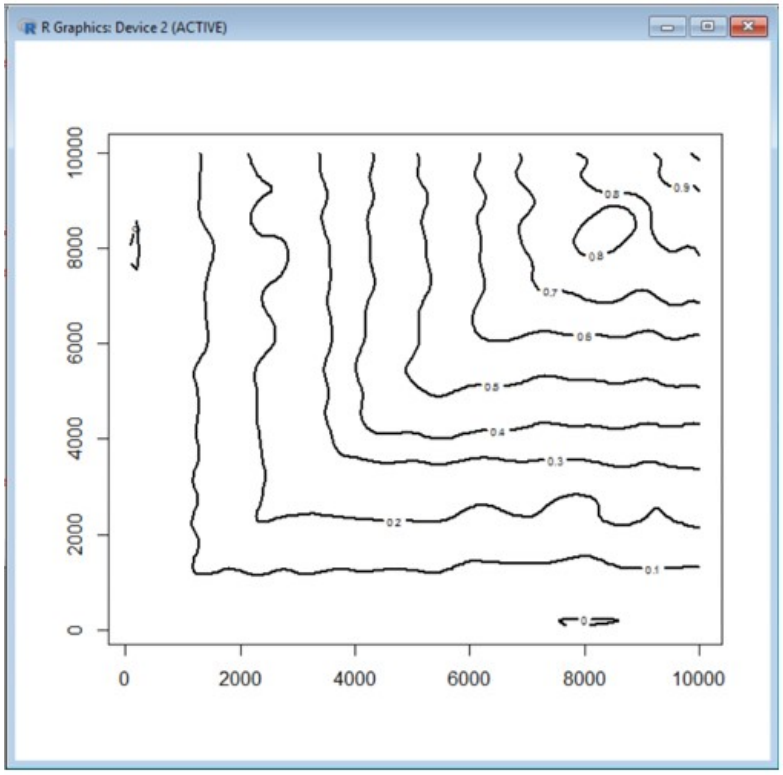


Figura 17: Un diagrama de contorno de la covarianza función de la muestra de 50 caminatas aleatorias Adaptación de Kokoszka, P. Reimherr, M .2017

5.6. LOS COMPONENTES PRINCIPALES FUNCIONALES ESTIMADOS (EFPC):

De acuerdo con Kokoszka, P. y Reimherr, M. (2017), Una de las herramientas más útiles y de uso frecuente en el análisis de datos funcionales, es el análisis de los componentes principales. Los componentes principales funcionales estimados, (EFPC) están relacionados con la función de covarianza muestral $\hat{c}(t,s)$.

En la expansión de bases:

$$X_n(t) \approx \sum_{m=1}^{M} c_{nm} B_m(t), \qquad 1 \leq n \leq N \tag{24}$$

las funciones básicas $B_m(t)$ son fijas. La idea de la expansión de componentes principales funcionales es encontrar funciones $\hat{v}_j$ de modo que el las funciones centradas $X_n - \bar{X}_N$ se representan como:

$$X_n(t) - \hat{X}_N(t) \approx \sum_{j=1}^{p} \hat{\xi}_{nj} \hat{v}_j(t) \tag{25}$$

con p mucho más pequeño que M en la ecuación anterior. Los $\hat{v}_j$ de los EFPC se calculan a partir de las funciones observadas $X_1, X_2, \ldots, X_N$ después de convertirlos a objetos funcionales. La Figura 18 muestra los $\hat{v}_j$ de los EFPC, para $j = 1, 2, 3, 4$; calculado para las 50 trayectorias suavizadas de la caminata aleatoria. Los $\hat{v}_j$ se asemeja a funciones trigonométricas. El primer EFPC, $\hat{v}_1$, trazado como la línea negra continua, muestra El patrón más pronunciado de la desviación de la función media de una trayectoria seleccionada al azar. Un examen superficial de la figura 11 confirma que la forma de $\hat{v}_1$ realmente resume bien el patrón principal de variabilidad alrededor la función media. Para cada curva X_n , el coeficiente $\hat{\xi}_{n1}$ cuantifica la contribución de $\hat{v}_1$ a su forma. El coeficiente $\hat{\xi}_{nj}$ se llama puntaje X_n de con respecto a $\hat{v}_1$. El segundo EFPC, trazado como la línea roja discontinua, muestra el segundo modo más importante de desviación de las funciones medias de las 50 curvas de caminatas aleatorias. Los $\hat{v}_j$ de EFPC son ortonormales, en el sentido de:

$$\int \hat{v}_j(t)\, \hat{v}_i(t)\, dt \begin{cases} 0 \ si \ j \ \neq i \\ 1 \ si \ j = i \end{cases} \tag{26}$$

Esta es una propiedad universal de los EFPC que restringe su capacidad de interpretación. Por ejemplo, $\hat{v}_2$ es el segundo modo más importante de variabilidad que es ortogonal a $\hat{v}_1$. Se que la variabilidad total de una muestra de curvas sobre la función media de la muestra, se puede afirmar que la variabilidad total de una muestra de curvas en torno a la función media muestral, puede descomponerse en la suma de variabilidades explicadas por cada EFPC. Para la muestra de las 50 caminatas aleatorias, la primera EFPC $\hat{v}_1$, explica aproximadamente el 81 % de la variabilidad, el segundo aproximadamente el 10 %, el tercero alrededor del 4 % y el cuarto alrededor del 2 %. Juntos, los primeros cuatro EFPC explican más del 96 % de variabilidad. Esto justifica la expansión usando p = 4, o incluso p = 2, como la contribución de los componentes restantes a la forma de las curvas es pequeño. El porcentaje de la variabilidad explicada por $\hat{v}_j$ está relacionado con el tamaño de las puntuaciones $\hat{\xi}_{nj}$; cuanto menor es el porcentaje, menores son los puntajes.

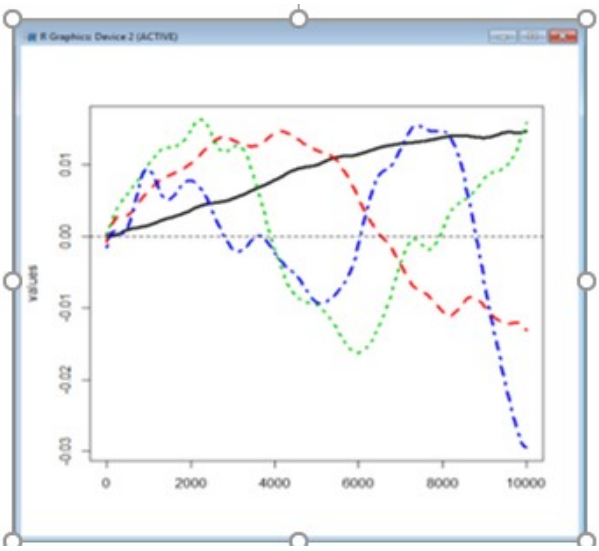


Figura 18: Los primeros cuatro componentes principales funcionales estimados de las 50 caminatas aleatorias. Adaptación de Kokoszka, P. Reimherr, M .2017

6. PRUEBA I DE MORAN PARA DATOS FUNCIONALES

De acuerdo a Balzanella y otros (2017) en su artículo se presenta una nueva estrategia para controlar la dependencia espacial a lo largo del tiempo, que adapta el método clásico del índice I de Moran para el control de la dependencia espacial al reto del tratamiento de flujos de datos funcionales o curvas de os datos en función del tiempo. Teniendo en cuenta que se puede determinar el índice I de Moran para datos funcionales mediante la siguiente expresión:

$$I = \frac{n}{\sum_i \sum_{i\prime} \ a_{i,i\prime}} \frac{\sum_i \sum_{i\prime} \ a_{i,i\prime} \int_0^T \left[\left[\bar{Y}_i^k(t) - \bar{Y}(t)\right]\left[\bar{Y}_{i\prime}^k(t) - \bar{Y}(t)\right]\right] dt}{\sum_i \int_0^T \left[\bar{Y}_i^k(t) - \bar{Y}(t)\right]^2 dt} \tag{27}$$

Donde:

n es el número de datos en cada curva estudiada.

i *e* i' son dos ubicaciones diferentes.

$\bar{Y}(t)$ es la curva media de las observaciones.

$\bar{Y}_i^k(t)$ es la curva de las observaciones en el lugar i.

$\bar{Y}_{i\prime}^k(t)$ es la curva de las observaciones en el lugar i'.

$\sum_i \sum_{i\prime} \ a_{i,i\prime} \int_0^T \left[\left[\bar{Y}_i^k(t) - \bar{Y}(t)\right]\left[\bar{Y}_{i\prime}^k(t) - \bar{Y}(t)\right]\right]$ Es la suma de la matriz del producto punto entre la resta de coeficientes de las observaciones y los coeficientes de las Medias.

$\sum_i \int_0^T \left[\bar{Y}_i^k(t) - \bar{Y}(t)\right]^2 dt$ Es la suma de la matriz al cuadrado de la resta de coeficientes de las observaciones y los coeficientes de las Medias.

$\sum_i \sum_{i\prime} a_{i,i\prime}$ corresponde a la doble sumatoria de las matrices de vecindad o de cercanías.

Para calcular la doble sumatoria de las matrices de vecindad o de cercanía se deben realizar los siguientes pasos:

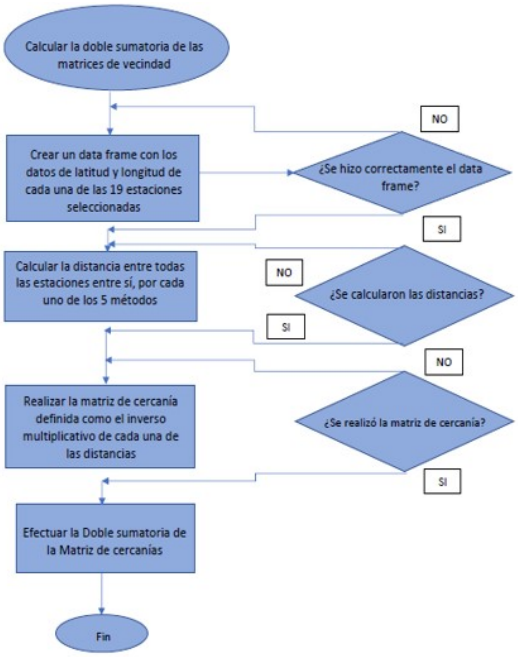


Figura 19: Diagrama de flujo para calcular la doble sumatoria de la matriz de vecindad o cercanía. Elaboración propia.

1. Crear un data frame con los datos de latitud y longitud de cada una de las 19 estaciones seleccionadas
2. Calcular la distancia entre todas las estaciones entre sí, por cada uno de los 5 métodos
3. Realizar la matriz de cercanía definida como el inverso multiplicativo de cada una de las distancias
4. Efectuar la Doble sumatoria de la Matriz de cercanías.

Para calcular el Índice I de Moran se deben realizar los siguientes pasos:

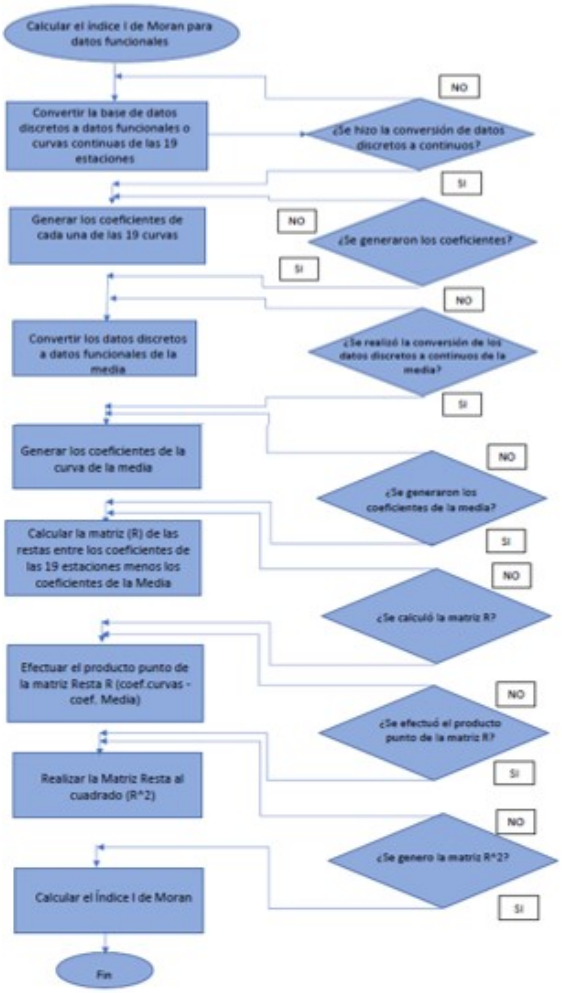


Figura 20: Diagrama de flujo para calcular el índice I de Morán para datos funcionales. Elaboración propia.

1. Convertir la base de datos discretos a datos funcionales o curvas continuas de las 19 estaciones.
2. Generar los coeficientes de cada una de las 19 curvas.
3. Convertir los datos discretos de la media a datos funcionales.
4. Generar los coeficientes de la curva de la media.

5. Calcular la matriz (R) de las restas entre los coeficientes de las 19 estaciones menos los coeficientes de la Media
6. Efectuar el producto punto de la matriz Resta R (coef.curvas - coef. Media)
7. Realizar la Matriz Resta al cuadrado (R^2)
8. Calcular el Índice I de Moran

Continuando con Balzanella y otros (2017), es necesario introducir nuevos enfoques para medir la autocorrelación espacial entre bases de datos discretas, ya que las principales medidas espaciales requieren un gran esfuerzo computacional, por esto se propone realizar la conversión a datos funcionales, para trabajar los datos como curvas continuas y calcular la autocorrelación espacial a partir de ellas y no con los datos originales.

7. APLICACÍON AL CONJUNTO DE DATOS

7.1. DEPURACION DEL CONJUNTO DE DATOS TRABAJADA

La base de datos inicial se obtuvo del Instituto de Hidrología, Meteorología y estudios ambientales "IDEAM", que es de dominio público y se actualiza mensualmente. Se puede observar en el link: `https://datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Precipitaci-n/s54a-sgyg/about_data`

Consta de 12 variables o columnas y aproximadamente 194´000.000 registros o filas, las variables son:

	Nombre	Descripción
1	CodigoEstacion	Es el número de identificación de la estación dentro del catálogo de estaciones
2	CodigoSensor	Código asignado al sensor
3	FechaObservacion	Fecha cuando se realizó la observación
4	ValorObservado	Valor medido en mm
5	NombreEstacion	Es el nombre dado a la estación dentro del catálogo de estaciones
6	Departamento	Nombre del departamento donde está ubicada la estación
7	Municipio	Nombre del municipio donde está ubicada la estación
8	ZonaHidrografica	Nombre de la zona Hidrográfica donde se ubica la estación
9	Latitud	Valor de la Latitud de la ubicación de la estación
10	Longitud	Valor de la Longitud de la ubicación de la estación
11	DescripcionSensor	Tipo de sensor, en esta base es de Precipitación
12	UnidadMedida	Milímetros (1 mm de lluvia equivale a 1 Lt de agua sobre una superficie de 1 m^2)

Tabla 2: Descripción de las variables de la base de datos

De esta base de datos inicial, se seleccionaron las correspondientes al municipio de Bogotá, dando como resultado una base de 19´114.556 registros o filas y 91 variables, correspondientes a las estaciones de pluviometría del IDEAM instaladas en la capital.

Con la información de cada estación se procedió a filtrar y unir las variables de interés: FechaObservacion y ValorObservado y debido a la gran cantidad de datos y para facilitar el análisis temporal, se sumó por días los datos del ValorObservado. A continuación, se armó una base de datos con la columna de fechas y las 91 estaciones como variables, siendo los registros el valor observado del pluviómetro por día, desde el 11/09/2018 hasta el 24/01/2022.

7.2. Relación de las 91 estaciones pluviométricas del IDEAM en Bogotá y sus ubicaciones

	Nombre	Latitud	Longitud	Altitud
1	ALTOS DE LA ESTANCIA - AUT [2120000101]	4,58070	-74,1783	2.736
2	ALTOS DE LA ESTANCIA - AUT [2120000101]	4.50000	-74,18	2.736
3	AMARILLOS LOS [35020440]	4,76670	-74,2167	3.500
4	Ministerio de Relaciones Exteriores de Arrayan-San [21200080]	4,58330	-74,0333	3.047
5	AUSTRALIA [21201300]	4,39425	-74.132	3.050
6	AY SAN FRANCISCO [21200120]	4,58330	-74,0333	3.000
7	BETANIA [35020350]	4,21890	-74,1467	3.150
8	BOCA GRANDE [21200190]	4,33330	-74,1333	3.460
9	CASABLANCA [21201970]	4,56670	-74,1667	2.665

10	CERRO DE SUBA [21200310]	4.75000	-740.667	2.691
11	CERRO NORTE - AUT [2120000103]	4,73260	-74,0187	2.766
12	CERRO NORTE - AUT [2120000103]	4.73000	-74,02	2.766
13	CORONEL INEM KENNEDY	4,62460	-74.155	2.557
14	CORONEL RAFAEL URIBE	4,58320	-74,1325	2.564
15	COLINA SORRENTO	4,61960	-74,1132	2.561
16	CORONEL ALBERTO LLERAS	4,74300	-74,1014	2.556
17	COL. ALEMANIA UNIFICADA	4,55580	-740.967	2.743
18	CORONEL ANTONIO GARCÍA	4,54500	-74,1394	2.602
19	COL. CIUDAD DE VILLAVICENCIO	4,48710	-74,1021	2.923
20	CORONEL EDUARDO UMAÑA	4.49000	-74,1131	2.840
21	COL. EL MANATIAL	4,56140	-74,0748	2.877
22	COL. EL TESORO DE LA CUMBRE	4,53530	-74,1491	2.816
23	CORONEL FRIEDRICH NAUMANN	4,74940	-74,0229	2.601
24	CORONEL GABRIEL GARCÍA MÁRQUEZ	4,50450	-74,0924	3.055
25	CORONEL GUSTAVO MORALES	4,71660	-74,0644	2.553
26	CORONEL GUSTAVO RESTREPO	4,57590	-741.062	2.576
27	COL. LA CHUCUA	4,60360	-74,1485	2.562
28	COL. LOS ALPES SEDE A	4,55260	-74,0837	2.875
29	COL. LOS PINOS SEDE A	4,58780	-74,0681	2.779
30	CORONEL MANUEL ELKIN PATARROYO	4,61820	-74,0644	2.672
31	COL. NUEVA ZELANDIA	4,76590	-74,0468	2.576
32	CORONEL OFELIA URIBE ACOSTA	4,50640	-74,1025	2.815
33	CORONEL PAULO FRIRE	4,53280	-74,1168	2.605
34	COLEGIO ALEMANIA SOLIDARIA - AUT [2120000100]	4,65240	-74,0754	2.556
35	COLEGIO ALEMANIA SOLIDARIA - AUT [2120000100]	4.65000	-74,08	2.556
36	COLEGIO CARLOS PIZARRO - AUT [2120000102]	4,63540	-74,2054	2.544
37	COLEGIO CARLOS PIZARRO - AUT [2120000102]	4.64000	-74,21	2.544
38	COLEGIO MIGUEL ANTONIO CARO - AUT [2120000104]	4,81320	-74,0313	2.575
39	COLEGIO MIGUEL ANTONIO CARO - AUT [2120000104]	4.81000	-74,03	2.575
40	COLEGIO RODOLFO LLINAS - AUT [2120000109]	4,72010	-74,1083	2.554
41	COLEGIO RODOLFO LLINAS - AUT [2120000109]	4.72000	-74,11	2.554
42	COLEGIO VEINTIUN ÁNGELES - AUT [2120000099]	4,74920	-74,0807	2.614
43	COLEGIO VEINTIUN ÁNGELES - AUT [2120000099]	4.75000	-74,08	2.614
44	CONTADOR [21200650]	4.70000	-74,0333	2.597
45	DELIRIO [21200130]	4,55000	-74,05	3.000
46	DIAMANTE [21190130]	4,01670	-74,2667	3.890
47	EL CÓDITO - AUT [2120000105]	4,76350	-74,0218	2.730
48	EL CÓDITO - AUT [2120000105]	4.76000	-74,02	2.730
49	ENMANUEL D ALZON [21201230]	4,70110	-74,0703	2.520
50	ESC. PEDAGÓGICA EXPRIMENTAL	4,67550	-74,0193	2.905
51	ESCUELA LA UNIÓN [21201200]	4,34290	-74,1839	3.320
52	ACUEDUCTO DE FONTIBON [21202120]	4,66670	-74,15	2.545
53	FUNDACIÓN ANA RESTREPO	4,70550	-74,0226	2.689
54	GALLO EL [35020430]	4,06670	-74,1333	3.440
55	GRAN BRETAÑA - AUT [2120000106]	4,51170	-74,1635	3.087
56	GRAN BRETAÑA - AUT [2120000106]	4.10000	-74,16	3.087
57	GRANIZO [21200320]	4,61670	-74,05	3.125
58	GUAMO EL [21190140]	3,96670	-74,2833	3.870

59	HATO EL [21200200]	4,38330	-74,1833	3.150
60	HORMONA-LAB [21200580]	4,68330	-74,0667	2.592
61	IDIGER - AUT [2120000108]	4.67000	-74,11	2.555
62	IDIGER - AUT [2120000108]	4,67490	-74,1136	2.555
63	ISLA LA [21202090]	4,63330	-74,2167	2.537
64	JARDÍN BOTÁNICO [21200610]	4,66670	-74,1	2.586
65	JUAN REY - AUT [21202190]	4,51670	-74,0833	2.682
66	LA FISCALIA - AUT [2120000107]	4,53600	-74,1065	2.718
67	LA FISCALA [2120000107]	4.54000	-74,11	2.718
68	LAG CHISACA [21200250]	4.30000	-74,2	3.770
69	MORALBA	4,54290	-74,0803	2.971
70	NAZARET [35020470]	4,16670	-74,15	2.600
71	PASQUILLA [21201580]	4,44650	-74,1548	3.000
72	POZO LLANITOS [35020410]	4,01670	-74,1333	2.850
73	REGADERA 2 LA [21200340]	4.40000	-74,15	3.056
74	RÍO NAZARET [21201350]	4,16670	-74,15	2.600
75	SAN ANTONIO [21190280]	4.10000	-74,3333	2.800
76	SAN DIEGO [21200230]	4,61670	-74,0667	2.700
77	SAN JUAN [21190270]	4,03100	-74,3112	2.900
78	SANTA MARÍA DE USME [21201240]	4,48130	-74,1263	2.800
79	SALSA 2 [21202070]	4.65000	-74,15	2.900
80	SEDE IDEAM CALLE 25D KRA [21202280]	4,68400	-74.129	2.589
81	SEMINARIO CONCILIA [21200040]	4.60000	-740.667	2.600
82	SIERRA MORENA - AUT [21202170]	4,56670	-74,1667	2.682
83	STA LUCIA [21200520]	4,56670	-74,1167	2.630
84	STA ROSA [21201290]	4,23330	-74,1833	3.430
85	TAQUES LOS [35025070]	4,19670	-74,1909	3.150
86	TORCA [21200770]	4,78330	-74,0333	2.579
87	TORQUITA [35020450]	4.01700	-74,2167	3.999
88	TUNAL EL CANDELARI [21200590]	4,56670	-74,15	2.599
89	USAQUEN [21201110]	4,68330	-740.167	2.647
90	VERJON EL [21200240]	4,58330	-740.167	3.250
91	VIEJA LA [21200660]	4.65000	-74,05	2.720

Tabla 3: Ubicación espacial de las 91 estaciones pluviométricas del IDEAM en Bogotá. Elaboración propia

Finalmente, se filtraron las estaciones que tenían la mayor cantidad de datos registrados, luego se imputaron los datos faltantes mediante el método de NNI (Nearest Neighbor Imputation o imputación por vecino más próximo) propuesto por Cohen (1996) al usar la variabilidad de los datos muestrales, en el método de imputación de la media, para agregar variabilidad a los valores imputados. Para finalmente llegar concretar una base de datos definitiva en este estudio, con 19 variables correspondientes a las estaciones seleccionadas, las cuales son: Estancia, Artillería, CerroNorte, Alemania, CarlosPizarro, MiguelAntonioCaro, RodolfoLLinas, Ángeles, ElCodito, ElDorado, GranBretaña, IDEAM Bogotá, IDIGER, JardínBotánico, LaFiscala, NuevaGeneración, SanFrancisco, UniversidadNacional y VillaTeresa y 1232 registros diarios, aproximadamente 3 años y 4 meses.

Ubicación de las 19 estaciones en el mapa de Bogotá	
1. Estancia	11. Gran Bretaña
2. Artillería	12. IDEAM Bogotá
3. Cerro Norte	13. IDIGER
4. Alemania	14. Jardín Botánico
5. CarlosPizarro	15. La Fiscala
6. Miguel Antonio Caro	16. Nueva Generación
7. Rodolfo Llinás	17. San Francisco
8. 21 Ángeles	18. Universidad Nacional
9. El codito	19. Villa Teresa
10. El Dorado	

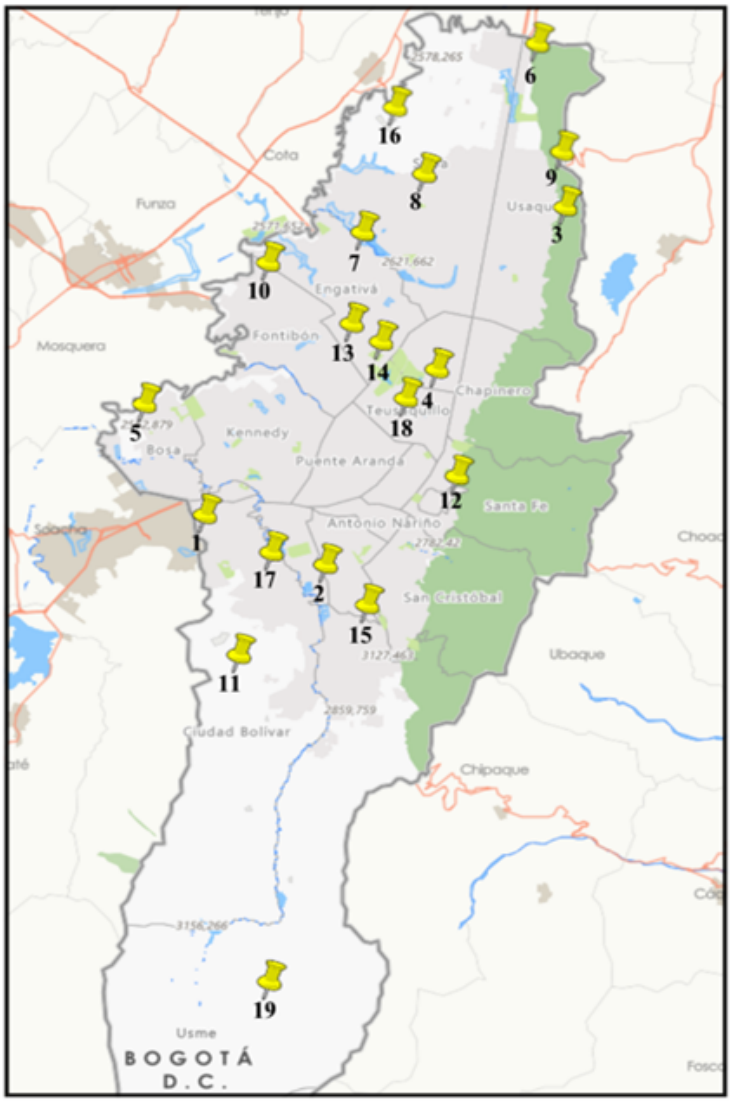


Figura 21: Listado y ubicación de las 19 estaciones pluviométricas seleccionadas de Bogotá. Elaboración propia

Por otro lado, se elaboró una base de datos con las ubicaciones geográficas de las 19 estaciones mencionadas anteriormente, a partir de las variables iniciales de la Latitud y la Longitud de la base de datos del IDEAM que recibió el nombre de coordenadas 1

	Latitud	**Longitud**
Estancia	4.581	-74.178
Artillería	4.556	-74.125
CerroNorte	4.733	-74.019
Alemania	4.652	-74.075
CarlosPizarro	4.635	-74.205
MiguelACaro	4.813	-74.031
RodolfoLLinas	4.720	-74.108
21Ángeles	4.749	-74.081
ElCodito	4.763	-74.022
ElDorado	4.706	-74.151
GranBretaña	4.512	-74.164
IDEAMB.	4.600	-74.067
IDIGER	4.675	-74.114
J.Botánico	4.669	-74.103
LaFiscala	4.536	-74.106
N.Generación	4.782	-74.094
S.Francisco	4.562	-74.149
U.Nacional	4.638	-74.089
VillaTeresa	4.350	-74.150

Tabla 4: Coordenadas geográficas de las 19 estaciones pluviométricas implicadas en el estudio. Elaboración propia

En la tabla N. 5 se pueden observar la comparación de los valores mínimos, 1° Cuartil, Mediana, Media, 3° cuartil, máximos y NA de las 19 estaciones antes y después de ser imputadas.

Estación	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
Estancia	0.000	0.000	0.000	1.664	0.625	30.300	864
Estanciaim	0.000	0.000	0.000	1.792	0.6	30.3	0
ElDorado	0.000	0.000	0.100	2.624	2.100	85.300	31
ElDoradoim	0.000	0.000	0.135	2.577	2.022	85.300	0
Artillería	0.000	0.000	0.000	1.978	1.300	53.400	774
Artilleríaim	0.000	0.000	0.000	1.996	1.100	53.400	0
CerroNorte	0.000	0.000	0.100	3.406	2.475	53.800	766
CerroNorteim	0.000	0.000	0.100	3.276	2.100	53.800	0
Alemania	0.000	0.000	0.000	3.258	2.200	58.500	747
Alemaniaim	0.000	0.000	0.000	2.817	1.800	58.500	0
CarlosPizarro	0.000	0.000	0.000	1.375	0.700	33.100	772
CarlosPizarroim	0.000	0.000	0.000	1.432	0.700	33.100	0
MiguelA.Caro	0.000	0.000	0.100	2.274	1.800	35.200	780
MiguelA.Caroim	0.000	0.000	0.100	2.372	1.700	35.200	0
RodolfoLLinas	0.000	0.000	0.000	3.083	2.325	67.400	772
RodolfoLLinasim	0.000	0.000	0.000	2.802	2.025	67.400	0
Ángeles	0.000	0.000	0.100	3.015	2.250	63.100	740
Ángelesim	0.000	0.000	0.100	2.878	2.000	63.100	0
ElCodito	0.000	0.000	0.100	3.602	2.350	70.900	766
ElCoditoim	0.000	0.000	0.100	3.624	2.400	70.900	0
G.Bretaña	0.000	0.000	0.100	1.993	1.600	36.900	773
G.Bretañaim	0.000	0.000	0.100	1.906	1.500	36.900	0
IDEAMB.	0.000	0.000	0.100	2.440	2.200	40.600	357
IDEAMB.im	0.000	0.000	0.100	2.522	2.300	40.600	0
IDIGER	0.000	0.000	0.000	2.627	1.800	81.300	702
IDIGERim	0.000	0.000	0.000	2.471	2.100	81.300	0
J.Botánico	0.000	0.000	0.000	2.777	2.500	53.600	88
J.Botánicoim	0.000	0.000	0.000	2.831	2.500	53.600	0
LaFiscala	0.000	0.000	0.000	2.276	1.200	64.500	760
LaFiscalaim	0.000	0.000	0.000	2.337	1.200	64.500	0
N.Generación	0.000	0.000	0.000	0.928	0.100	31.000	579
N.Generaciónim	0.000	0.000	0.000	1.030	0.100	31.000	0
S.Francisco	0.000	0.000	0.000	1.640	0.925	53.700	772
S.Franciscoim	0.000	0.000	0.000	1.542	0.800	53.700	0
U.Nacional	0.000	0.000	0.000	1.780	0.900	41.500	514
U.Nacionalim	0.000	0.000	0.000	1.834	1.000	41.500	0
VillaTeresa	0.000	0.100	0.200	1.206	0.900	30.900	534
VillaTeresaim	0.000	0.100	0.200	1.197	0.900	30.900	0

Tabla 5: Comparativo entre los valores de las estaciones. iniciales y sus imputaciones. Elaboracion Propia

Se pueden apreciar a continuación, los Histogramas comparativos de las precipitaciones de la estación ElDorado antes (Figura 22) y después de ser imputada (Figura 23), donde no se evidencia un cambio relevante.

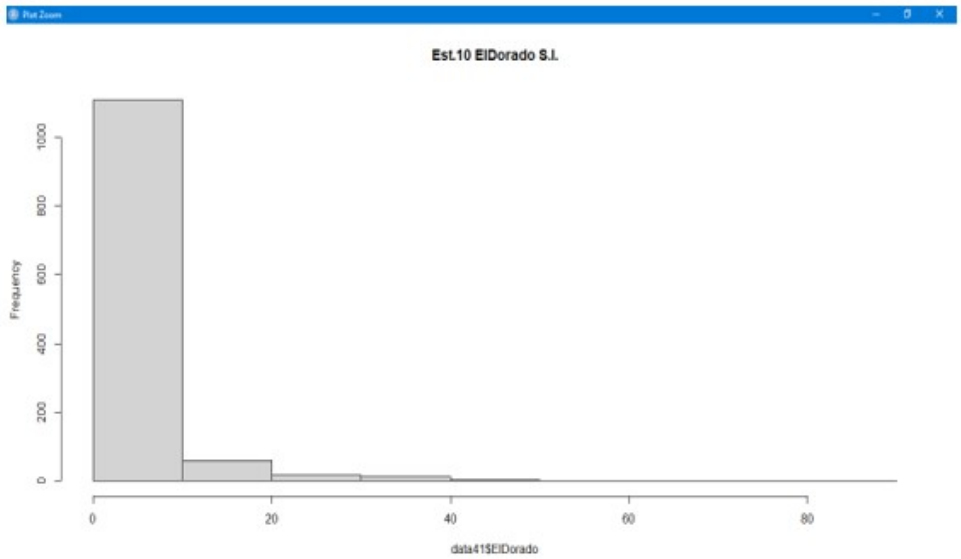


Figura 22: Histograma de los datos de la estación El Dorado antes de ser imputados. Elaboración propia

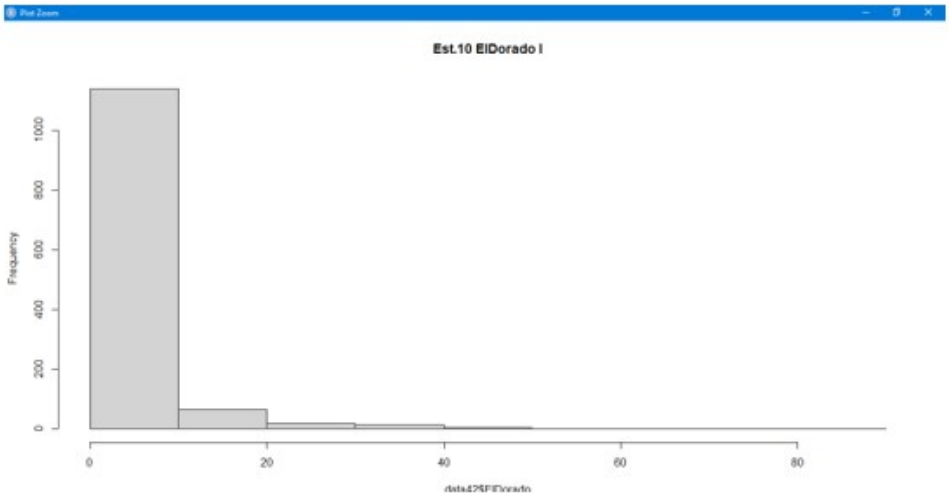


Figura 23: Histograma de los datos de la estación El Dorado después de ser imputados. Elaboración propia

En el anexo 2 se pueden observar los Histogramas comparativos de las precipitaciones en las 19 estaciones pluviométricas seleccionadas antes y después de ser imputadas.

Además, se presentan los Boxplots comparativos de las precipitaciones en la estación Artillería antes (Figura 24) y después de ser imputada (Figura 25), donde tampoco no se evidencia un cambio relevante.

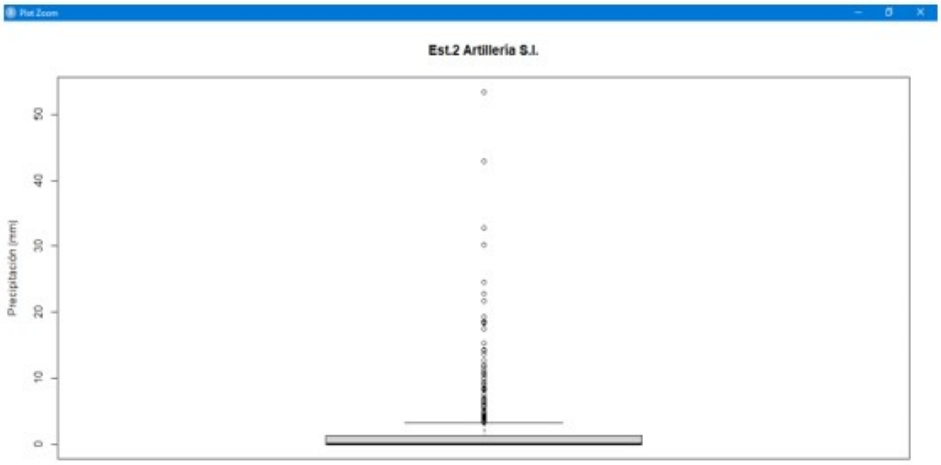


Figura 24: Boxplot de los datos de la estación Artillería antes de ser imputados. Elaboración propia

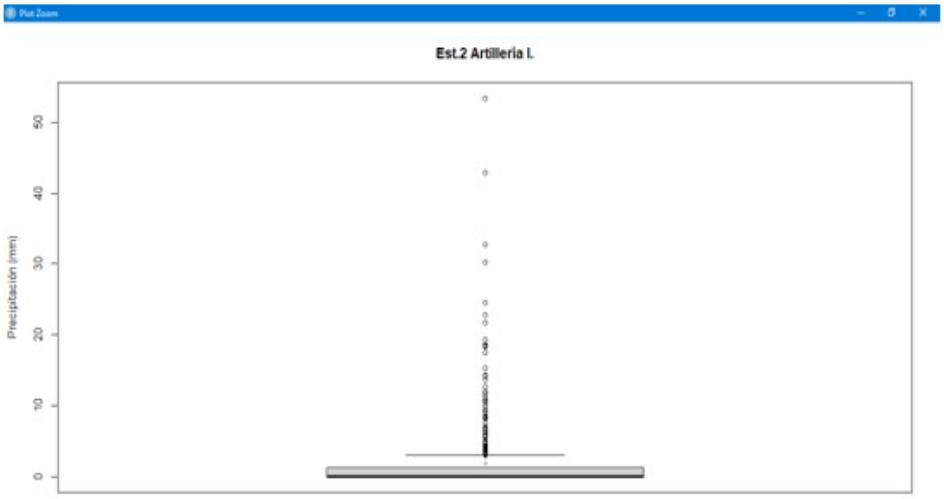


Figura 25: Boxplot de los datos de la estación Artillería después de ser imputados. Elaboración propia

En el anexo 3 se muestran los Boxplots comparativos de las precipitaciones en las 19 estaciones pluviométricas seleccionadas antes y después de ser imputadas.

A continuación, se muestran los datos discretos de la estación VillaTeresa de los 1232 días estudiados y la media en color azul, en la figura 26.

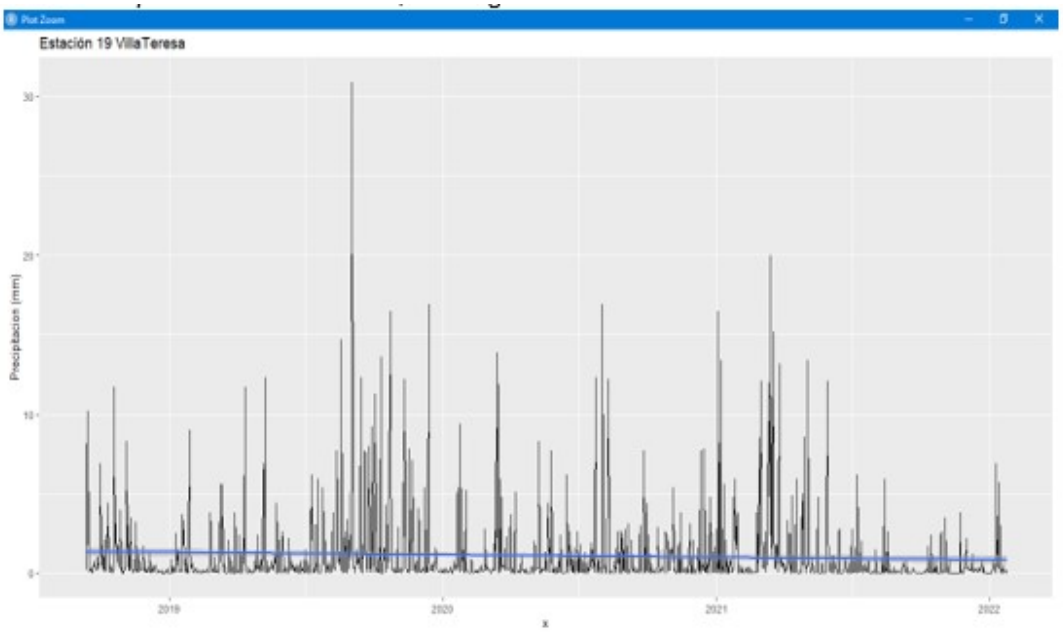


Figura 26: Visualización de los datos discretos de la estación VillaTeresa. Elaboración propia

En el anexo 1 se observan los datos discretos de la de las precipitaciones en las 19 estaciones pluviométricas seleccionadas.

Ya definido el conjunto de datos con los que se realiza este estudio, se pueden observar gráficamente, en la figura 27, el total de ellos en las 19 estaciones, durante los 1232 días de estudio, como datos discretos sin discriminar por estaciones así:

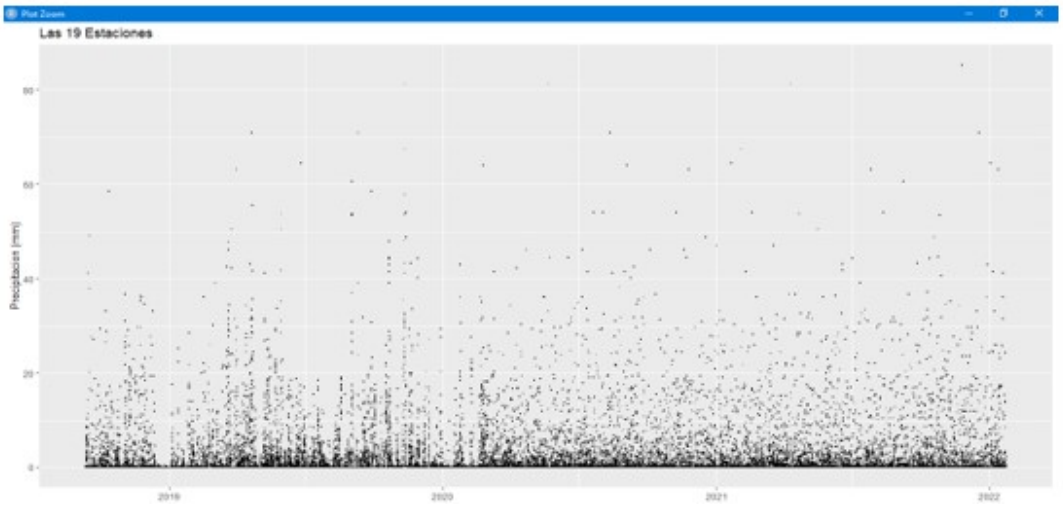


Figura 27: Visualización de los datos discretos de las 19 estaciones estudiadas sin discriminación. Elaboración propia

Y en la figura 28, se observan el total de ellos en las 19 estaciones, durante los 1232 días de estudio, como datos discretos discriminados por estaciones.

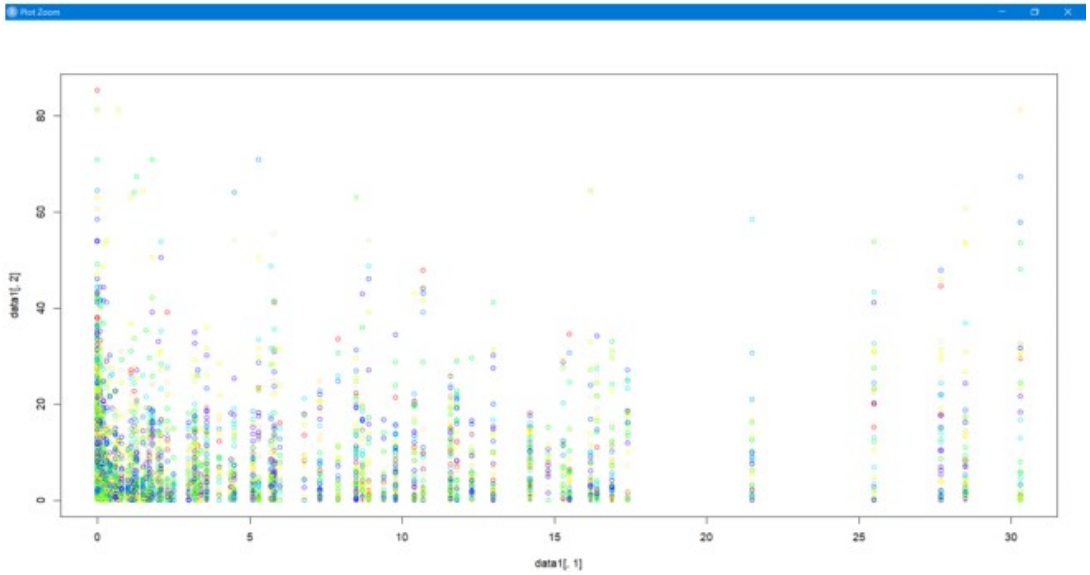


Figura 28: Visualización de los datos discretos de las 19 estaciones estudiadas discriminadas. Elaboración propia

A continuación, en las tablas 7 y 8 se pueden observar el encabezado de las primeras 11 observaciones de las medidas de precipitación de las 19 estaciones seleccionadas con los registros más completos.

Estancia	Dorado	Artillería	Cerro N.	Alemania	Carlos P.	Miguel A.C.	Rodolfo LL.	Ángeles
0	0	0	0,2	0	0	0	0	0
1,3	1,8	3,1	4,9	6,5	0	1,8	8,4	5,7
0	1,3	0,6	0	0	0	0,2	0	0
0	2,1	0	8,6	8,4	0	0,3	0	3,5
0,3	5,1	0	0	0	0	1,9	0,1	14,3
0	0	0	0,7	0,8	0	0	0	20,1
0	37,8	0	0	2,1	7,1	2,8	16,7	0,1
0,2	0	0	0,4	0	0	0	1,9	0,1
1,2	0	0	0	1,8	0	0	0	0
16,2	0	0	0,2	0	0	0	0	9,2
0	0	3,2	0	0,1	0	4,3	0,3	0

Tabla 6: Primeras 11 observaciones de pluviometría diaria en las primeras 9 estaciones estudiadas. Elaboración propia

Codito	Bretaña	IDEAMB	IDIGER	J.Botán.	Fiscala	N.Genrción	S.Franc.	U.Nacional	VillaTeresa
0	0	0	0	0	0	0	0	0	0,2
2,5	0	6,5	3,3	3,3	0,2	0	1,5	8,6	6,1
0	4,3	6,8	1,3	2,1	0	0	3,2	5,4	10,2
0,1	5,3	8,2	2,7	4,9	11,3	0	3,2	2,1	1,8
0,1	0	4,7	5	4,6	0	0	41,2	6,1	5,1
0	0	0,1	0	0	0	0	0,9	0	0,4
0	0	0,1	0	49,1	0	0	0	5,9	0,1
0	4,6	0	0	0	0	0	0	0,3	0,2
27,8	2,9	0	0	0	3,3	0	0	0,1	0,3
0	0	0	0	0	0	0	0	0	0
0	0	0,1	0	0	0,1	0	0	0,4	0,1

Tabla 7: Primeras 11 observaciones de pluviometría diaria en las segundas 10 estaciones estudiadas. Elaboración propia

7.3. EXPLORACIÓN GRÁFICA DE LOS OBJETOS FUNCIONALES O CURVAS DE LAS 19 ESTACIONES ESTUDIADAS:

En la figura 29 se observan las curvas de las precipitaciones registradas en las 19 estaciones seleccionadas, durante los 1.232 días, con la estación del Dorado que presenta los 2 picos más altos (alrededor de 5 mm),3 curvas tienen los valores más bajos, LaFiscala, UniversidadNacional y SanFrancisco, entre 0 y 1 mm; y las otras 15 curvas de las restantes estaciones tienen valores entre 1 y 4 mm.

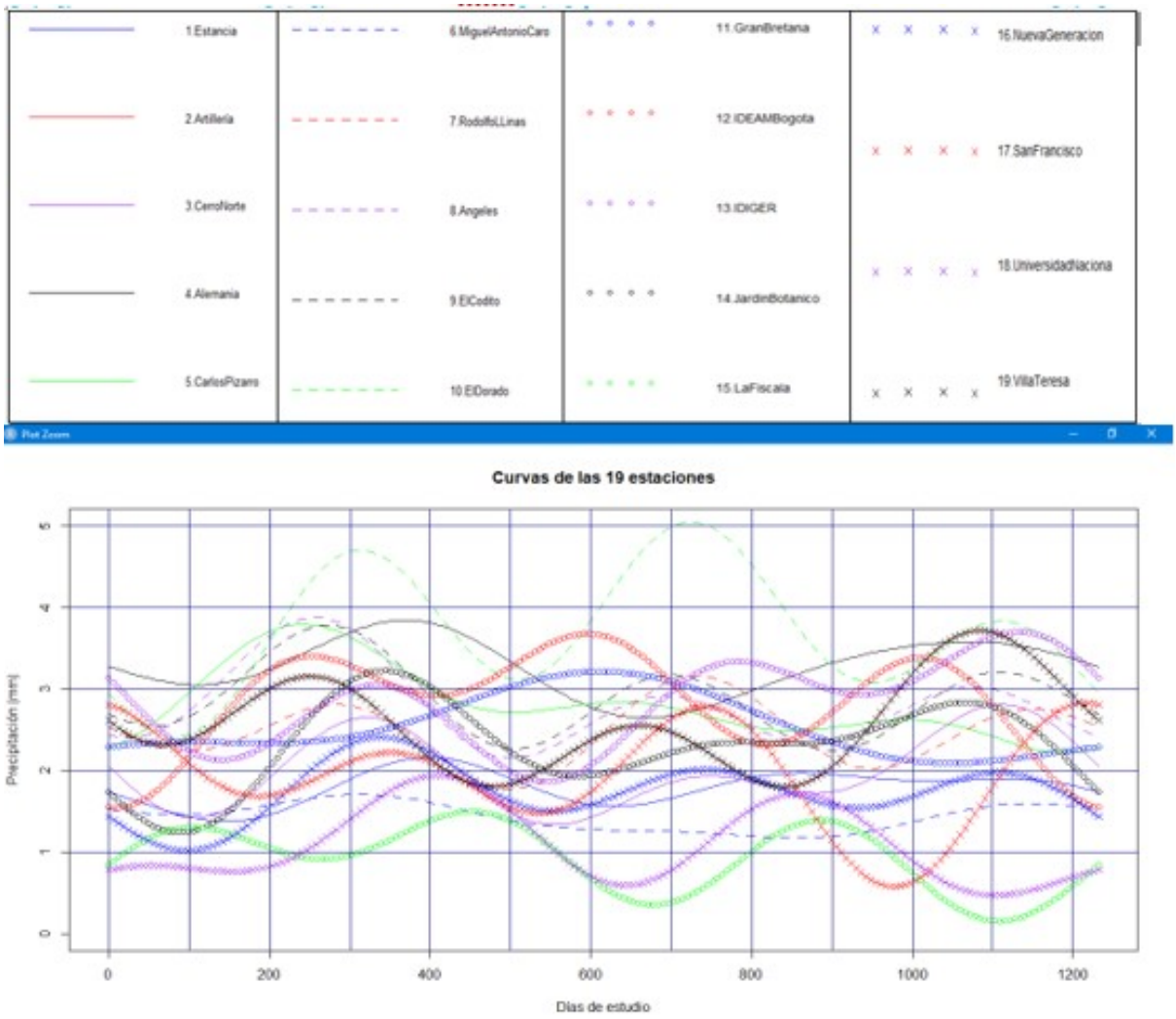


Figura 29: . Las curvas de lluvia de las 19 estaciones pluviométricas del IDEAM en Bogotá. Elaboración propia

En la figura 30 se observan el conjunto general de las curvas de las precipitaciones registradas en las 19 estaciones seleccionadas, durante los 1.232 días, con la media y el boxplot. Se puede apreciar que al inicio la media del periodo se presenta el valor de la media más baja ≈ 1, luego presenta un comportamiento constante ≈ 2 y al final toma su mayor valor ≈ 4

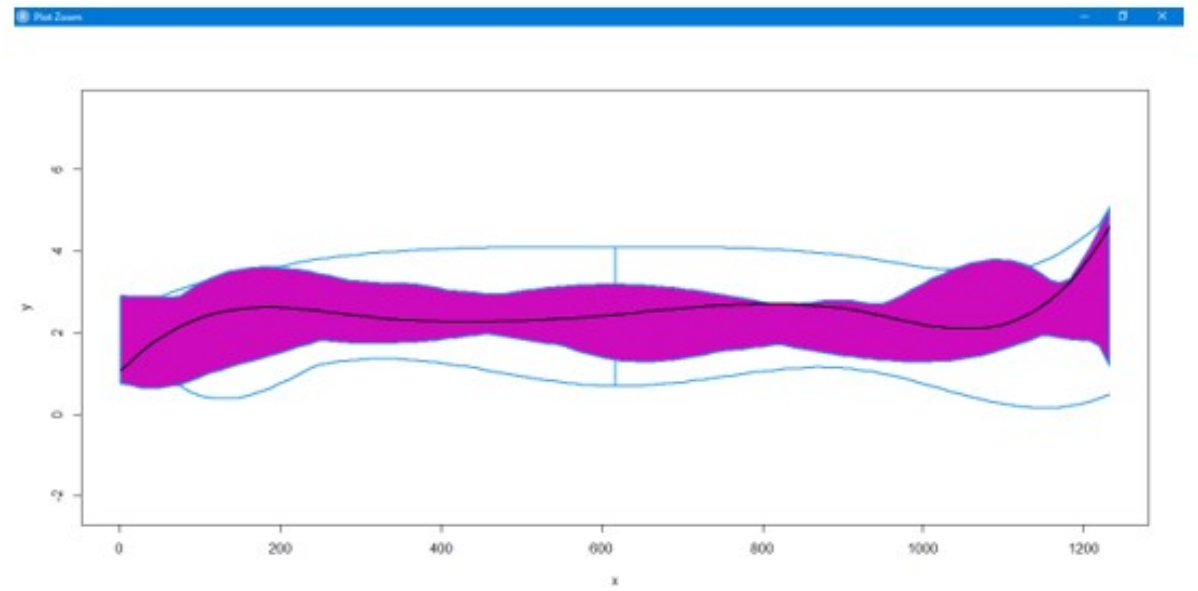


Figura 30: Las curvas de lluvia de las 19 estaciones pluviométricas del IDEAM en Bogotá, con la media y boxplot. Elaboración propia

En la figura 31 se observa el comportamiento de la media de las 19 estaciones seleccionadas durante los 1.232 días. Se puede apreciar que al inicio la media del periodo presenta su valor más bajo ≈ 1.82, en seguida se presenta un pico de ≈ 2.6 a los 350 días aproximadamente, luego un valle medio de ≈ 2.15 a los 650 días, a continuación, un pico medio ≈ 2.35 a los 950 días, otro valle pequeño de ≈ 2.3 a los 1.100 días y finalmente el valor máximo de ≈ 2.8 a los 1.200 días

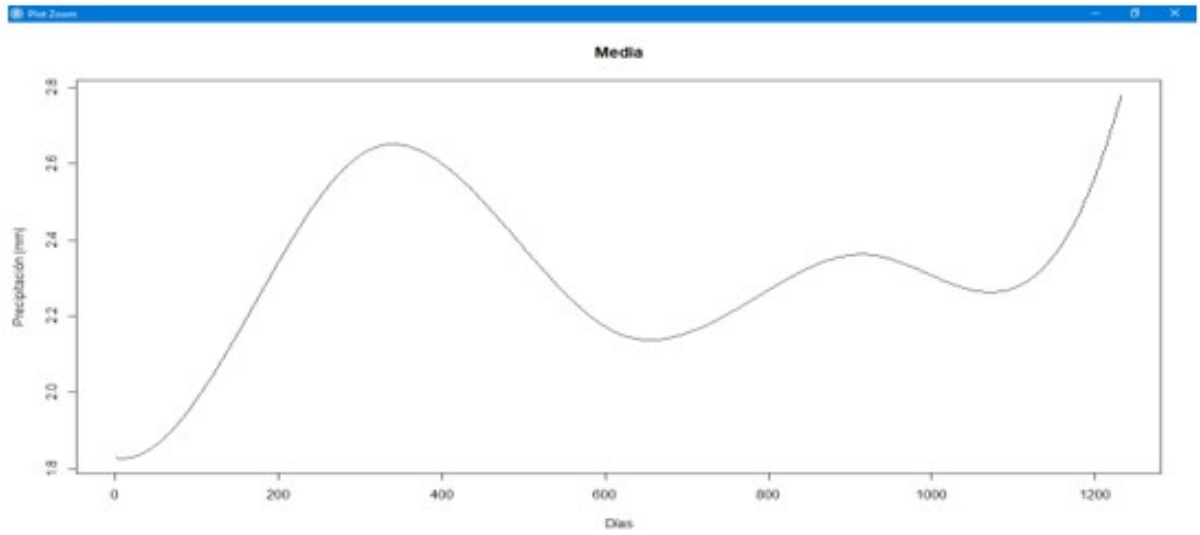


Figura 31: Media de las 19 estaciones pluviométricas del IDEAM en Bogotá. Elaboración propia

En la figura 32 se observa el comportamiento de la desviación estándar de las 19 estaciones seleccionadas durante los 1.232 días. Se puede apreciar que al inicio se presenta el valor más bajo≈ 0.5, en seguida se presenta un pico de ≈ 0.9 a los 150 días aproximadamente, luego un valle medio de ≈ 0.7 a los 350 días, a continuación, un pico medio ≈ 0.85 a los 650 días, otro valle pequeño de ≈ 0.7 a los 900 días y

finalmente el valor máximo de ≈ 1.1 a los 1.200 días

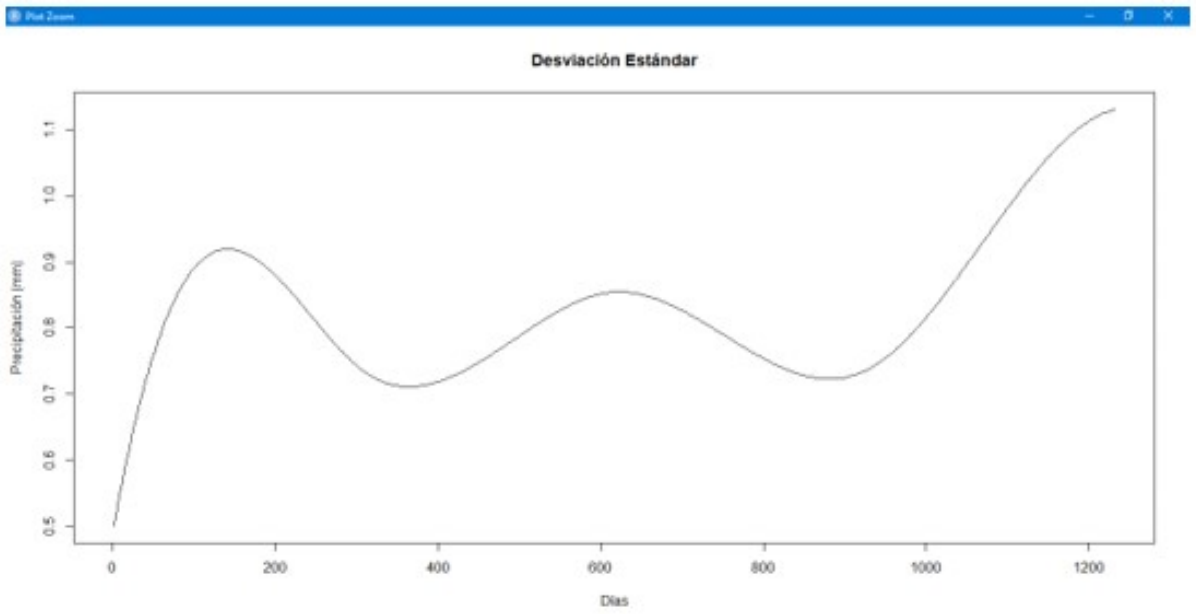


Figura 32: Desviación Estándar de las 19 estaciones pluviométricas del IDEAM en Bogotá. Elaboración propia

En la figura 33 se observa el diagrama de contorno de la covarianza de los registros de las 19 estaciones pluviométricas durante los 1.232 días de estudio, donde se aprecia que la covarianza es mayor en los valores de los días iniciales y finales.

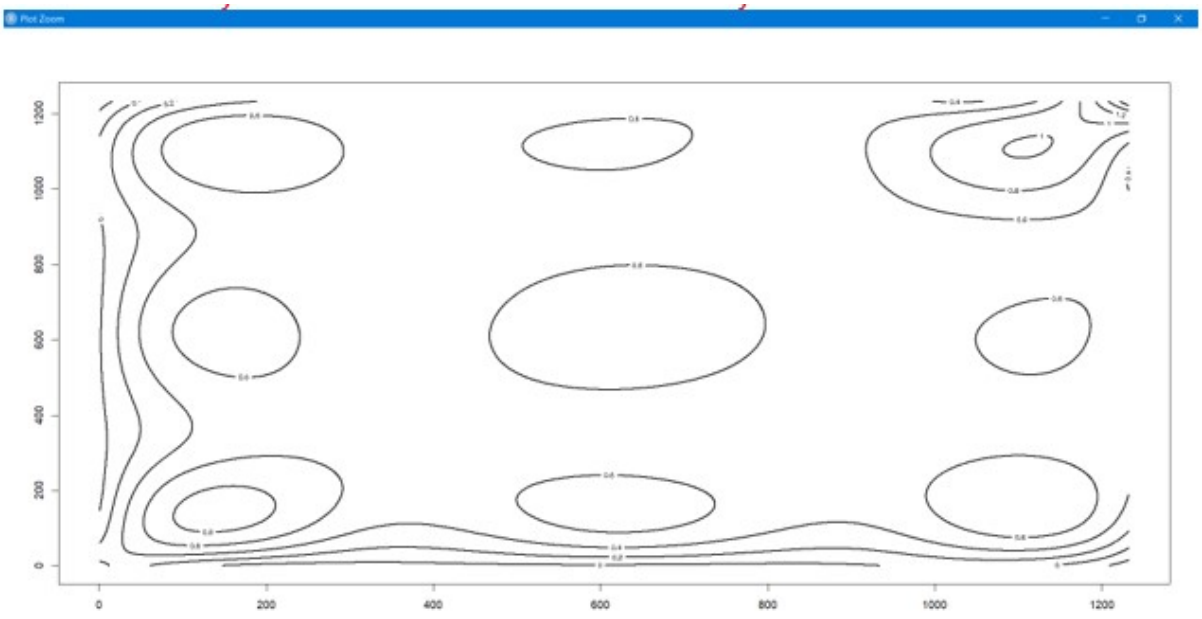


Figura 33: Diagrama de contorno de la covarianza de los registros de las 19 estaciones pluviométricas durante los 1232 días de estudio. Elaboración propia

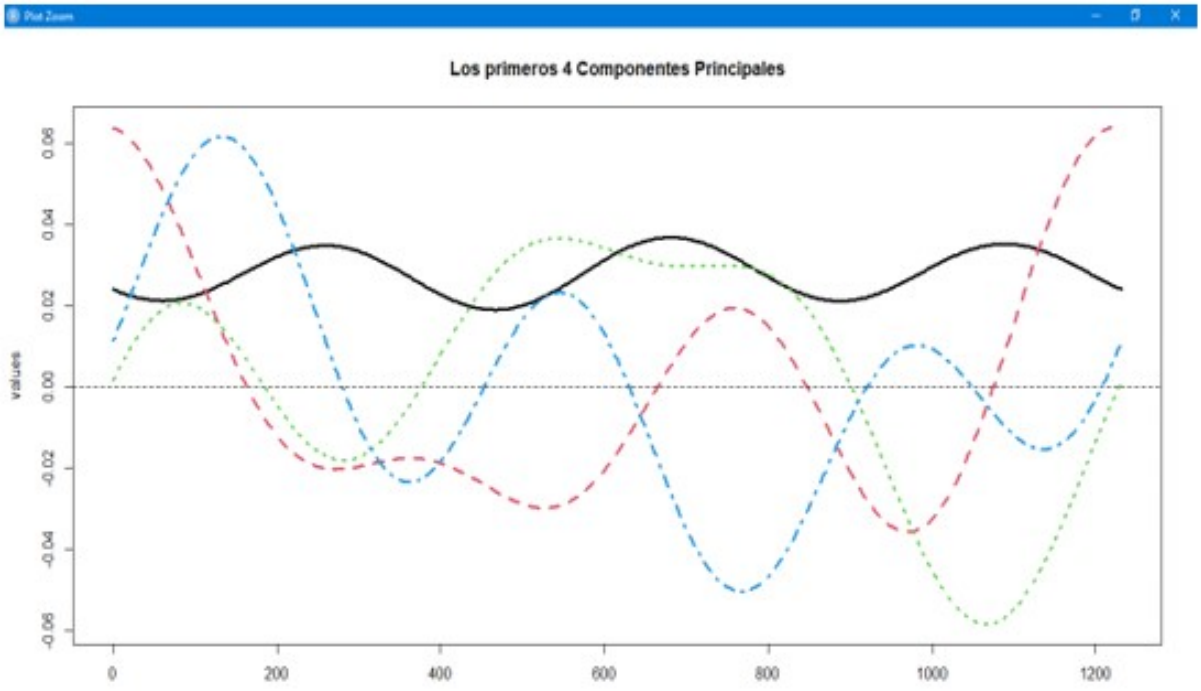


Figura 34: Los 4 Componentes Principales de las 19 estaciones pluviométricas durante los 1232 días de estudio. Elaboración propia

De acuerdo al gráfico anterior y según los cálculos realizados, el primer componente principal representa el 78,32 % de los datos, el segundo el 6,31 %, el tercero el 5,98 % y el cuarto el 4,24 %

8. CÁLCULO E INTERPRETACIÓN DE LA PRUEBA I DE MORAN PARA LOS DATOS DISCRETOS Y LOS FUNCIONALES.

8.1. CÁLCULO DEL ÍNDICE DE AUTOCORRELACIÓN I DE MORAN PARA LOS DATOS DISCRETOS:

Se trabajó la función Moran.I del paquete Ape, para las 6 tablas planteadas en el presente trabajo con los datos como variables discretas. Esta función calcula el coeficiente de autocorrelación I de Moran de x dando una matriz de pesos utilizando el método descrito por Gittleman y Kot (1990). La herramienta da como resultado los siguientes elementos: Valor Observado: la prueba I de Moran calculada Valor Esperado: el valor esperado de I bajo la hipótesis nula.

Sd: la desviación estándar de I bajo la hipótesis nula.

p.valor: el valor P de la prueba de la hipótesis nula frente a la hipótesis alternativa especificada en alternativa

N.	Tabla	Valor Observado	Valor Esperado	Sd	p.valor
1	Con las 19 estaciones seleccionadas	**No se pudo calcular. Aparece el error:** En mean.default (x): el argumento no es numérico ni lógico: devuelve NA			
2	Con 18 estaciones (Sin Villa Teresa)	0.0017	-4.6298e-5	9.7373e-5	0
3	Con las 3 estaciones del nororiente (CerroNorte, MiguelA.Caro y ElCodito)	0.0019	-0.00028	0.00049	1.5508e-05
4	Con las 4 estaciones centrales (Alemania, IDIGER, J.Botánico y U.Nacional)	0.0021	-0.00021	0.00038	8.6287e-10
5	Con las 3 estaciones del sur, (Artillería, LaFiscala y S. Francisco)	0.0018	-0.00028	0.00049	2.7583e-05
6	Con las 3 estaciones del occidente (CarlosPizarro, El Dorado y N.Generación)	0.0081	-0.00028	0.00049	0

Tabla 8: Resultados prueba Moran.I de los datos discretos para las 6 tablas planteadas. Elaboración propia

En todos los casos anteriores se puede observar que los datos manifiestan un ligero grado de dispersión en su distribución, pero se clasifican como datos sin dependencia espacial. Es necesario anotar que este método de cálculo de la prueba I de Moran para datos discretos, aunque en teoría no tiene límites respecto a la cantidad de datos presentes en las bases, manifiesta dificultades muy grandes cuando la base de datos es suficientemente extensa. Lo que da a entender que es más favorable trabajar bases de datos robustas mediante las técnicas de los datos funcionales.

8.2. TIPOS DE DISTANCIA UTILIZADOS PARA EL CÁLCULO DE LA MATRIZ DE CERCANÍA:

Con el fin de confirmar el resultado del cálculo del índice I de Moran, se proponen 5 métodos de medición de distancias para aplicar en las matrices de cercanías, en el conjunto de datos de las 19 estaciones pluviométricas en general y en los 5 subconjuntos planteados, en particular. Según Cuadras, C. (1989) se pueden resumir las características generales de los siguientes métodos de medir distancias así:

1. **Euclidiana Sexagesimal**

Es la distancia en línea recta entre 2 lugares a partir de sus coordenadas geográficas de latitud y longitud, suponiendo que se encuentran sobre un plano bidimensional.

2. **Euclidiana Métrica**
   Es la distancia en línea recta entre 2 lugares a partir de la conversión de sus coordenadas geográficas de latitud y longitud a metros y suponiendo que se encuentran sobre un plano bidimensional.

3. **Manhattan**
   Es la distancia entre 2 puntos ubicados en una cuadrícula, de tal manera que sólo pueden realizarse desplazamientos horizontales o verticales, es más realista que las distancias Euclidianas al tener en cuenta posibles obstáculos en la trayectoria en línea recta que trabajan estas.

4. **Minkowski**
   Es una métrica en un espacio vectorial normalizado, siendo un método de medir distancias más general, ya que dependiendo de un parámetro "p" puede variar la forma de la trayectoria. Se considera una generalización tanto de la distancia Euclidiana, como la de Manhattan.

5. **Chebysshev**
   Distancia definida sobre un espacio vectorial que calcula la distancia entre 2 puntos definidos por vectores, como la máxima diferencia a lo largo de cualquiera de sus dimensiones.

8.3. CÁLCULO DEL ÍNDICE DE AUTOCORRELACIÓN I DE MORAN PARA LOS DATOS FUNCIONALES:

A continuación, se muestran los valores del índice I de Moran para datos funcionales obtenidos según el tipo de distancia aplicada y de acuerdo a las diferentes tablas que se plantean, primero las 19 estaciones, después excluyendo la estación villa teresa que es la más distante, a continuación, el subgrupo de las 3 estaciones del nororiente (CerroNorte, MiguelACaro y ElCodito), luego el subgrupo de las 4 estaciones centrales (Alemania, IDIGER, JBotánico y UNacional.), seguida del subgrupo de las 3estaciones del sur, (Artillería, LaFiscala y SanFrancisco) y finalmente las 3 estaciones del occidente (CarlosPizarro, ElDorado y NGeneración) así:

N.	Tipo de distancia	DSMC	I
1	**Euclidiana Sexagesimal**	**3.514,702**	$4,65102 * 10^{-17} \approx 0$
2	**Euclidiana Métrica**	**3.514,702**	$4,65102 * 10^{-17} \approx 0$
3	**Manhattan**	**3.986,558**	$4,100517 * 10^{-17} \approx 0$
4	**Minkowski**	**3.741,17**	$4,369475 * 10^{-17} \approx 0$
5	**Chebysshev**	**3741,17**	$4,369475 * 10^{-17} \approx 0$

Tabla 9: Resultados de la prueba I de Moran según las 5 distancias para las 19 estaciones seleccionadas. Elaboración propia

N.	Tipo de distancia	DSMC	I
1	**Euclidiana Sexagesimal**	**3.389,99**	$-1.50403 * 10^{-17} \approx 0$
2	**Euclidiana Métrica**	**3.389,99**	$-1.50403 * 10^{-17} \approx 0$
3	**Manhattan**	**3.614,62**	$-1.410586 * 10^{-17} \approx 0$
4	**Minkowski**	**3.614,62**	$-1.410586 * 10^{-17} \approx 0$
5	**Chebysshev**	**3.859,61**	$-1.321046 * 10^{-17} \approx 0$

Tabla 10: Resultados de la prueba I de Moran según las 5 distancias para las 18 estaciones seleccionadas, descartando Villa Teresa. Elaboración propia

N.	Tipo de distancia	DSMC	I
1	**Euclidiana Sexagesimal**	**130,43**	$3,608803 * 10^{-16} \approx 0$
2	**Euclidiana Métrica**	**130,43**	$3,608803 * 10^{-16} \approx 0$
3	**Manhattan**	**116,24**	$4,049147 * 10^{-16} \approx 0$
4	**Minkowski**	**131,54**	$3,57835 * 10^{-16} \approx 0$
5	**Chebysshev**	**131,67**	$3,574871 * 10^{-16} \approx 0$

Tabla 11: . Resultados de la prueba I de Moran según las 5 distancias para las 3 estaciones seleccionadas del nororiente CERRO NORTE, MIGUEL ANTONIO CARO Y EL CODITO. Elaboracíon propia

N.	Tipo de distancia	DSMC	I
1	**Euclidiana Sexagesimal**	**193,15**	$1.630996 * 10^{-16} \approx 0$
2	**Euclidiana Métrica**	**193,15**	$1.630996 * 10^{-16} \approx 0$
3	**Manhattan**	**146,93**	$2.143981 * 10^{-16} \approx 0$
4	**Minkowski**	**207.77**	$1.516199 * 10^{-16} \approx 0$
5	**Chebysshev**	**229,85**	$1.370586 * 10^{-16} \approx 0$

Tabla 12: Resultados de la prueba I de Moran según las 5 distancias para las 3 estaciones seleccionadas del sur ARTILLERÍA, LAFISCALA Y SAN FRANCISCO. Elaboracíon propia

N.	Tipo de distancia	DSMC	I
1	Euclidiana Sexagesimal	467,88	$-1.8990 * 10^{-16} \approx 0$
2	Euclidiana Métrica	467,88	$-1.8990 * 10^{-16} \approx 0$
3	Manhattan	342,73	$-2.5925 * 10^{-16} \approx 0$
4	Minkowski	510,31	$-1.7411 * 10^{-16} \approx 0$
5	Chebysshev	560,23	$-1.5860 * 10^{-16} \approx 0$

Tabla 13: Resultados de la prueba I de Moran según las 5 distancias para las 4 estaciones seleccionadas del centro ALEMANIA, IDIGER, JARDÍN BOTÁNICO Y U.N. Elaboración propia

N.	Tipo de distancia	DSMC	I
1	**Euclidiana Sexagesimal**	**54,34**	**0**
2	**Euclidiana Métrica**	**54,34**	**0**
3	**Manhattan**	**38,81**	**0**
4	**Minkowski**	**60,41**	**0**
5	**Chebysshev**	**67,99**	**0**

Tabla 14: Resultados de la prueba I de Moran según las 5 distancias para las 3 estaciones seleccionadas del occidente CARLOS PIZARRO, EL DORADO Y NVA. GENERACIÓN. Elaboracíon propia

De acuerdo a los resultados observados en la tabla 10 donde para todos los 5 tipos de distancia, la prueba I de Moran para datos funcionales, da como resultado prácticamente 0, y según la teoría encontrada, (Spatio-Temporal Data. Cressi, N., Contemporary Statistical Models for the Plant and Soil Sciences. Schabenberger, O.), significa que no existe correlación espacial entre las 19 estaciones de pluviometría del IDEAM, ubicadas en Bogotá.

Resultado similar se presenta según la tabla 11 donde se descarta la estación Villa Teresa, que es la que está ubicada en el lugar más distante, el índice I de Moran da como resultado prácticamente 0, y según la teoría encontrada, (Spatio-Temporal Data. Cressi, N., Contemporary Statistical Models for the Plant and Soil Sciences. Schabenberger, O.), significa que no existe correlación espacial entre las 18 estaciones de pluviometría del IDEAM, ubicadas en la parte urbana de Bogotá.

Algo semejante se puede deducir de la tabla 12 donde el índice I de Moran da como resultado prácticamente 0, según las 5 distancias para las 3 estaciones seleccionadas del nororiente CERRO NORTE, MIGUEL ANTONIO CARO Y EL CODITO y según la teoría encontrada, (Spatio-Temporal Data. Cressi, N., Contemporary Statistical Models for the Plant and Soil Sciences. Schabenberger, O.), significa que no existe correlación espacial entre las 3 estaciones de pluviometría del IDEAM, ubicadas en el nor oriente de la ciudad de Bogotá.

Se sigue presentando un resultado similar en la tabla 13 Resultados de la prueba I de Moran según las 5 distancias para las 4 estaciones seleccionadas del centro de, ALEMANIA, IDIGER, JARDÍN BOTÁNICO

Y U. Nacional donde el índice I de Moran da como resultado prácticamente 0, según las 5 distancias y según la teoría encontrada, (Spatio-Temporal Data. Cressi, N., Contemporary Statistical Models for the Plant and Soil Sciences. Schabenberger, O.), significa que no existe correlación espacial entre las 4 estaciones de pluviometría del IDEAM, ubicadas en el centro de la ciudad de Bogotá.

De nuevo se puede deducir de la tabla 14 donde el índice I de Moran da como resultado prácticamente 0, según las 5 distancias para las 3 estaciones seleccionadas las 3 estaciones seleccionadas del sur ARTILLERÍA, LAFISCALA Y SAN FRANCISCO y según la teoría encontrada, (Spatio-Temporal Data. Cressi, N., Contemporary Statistical Models for the Plant and Soil Sciences. Schabenberger, O.), significa que no existe correlación espacial entre las 3 estaciones de pluviometría del IDEAM, ubicadas en el sur de la ciudad de Bogotá.

Finalmente se sigue corroborando el resultado dado en los subgrupos anteriores para las 3 estaciones seleccionadas del occidente de la ciudad: CARLOS PIZARRO, EL DORADO Y NVA. GENERACIÓN, tabla 15 donde el índice I de Moran da como resultado prácticamente 0, según las 5 distancias y según la teoría encontrada, (Spatio-Temporal Data. Cressi, N., Contemporary Statistical Models for the Plant and Soil Sciences. Schabenberger, O.), significa que no existe correlación espacial entre las 3 estaciones de pluviometría del IDEAM, ubicadas en el occidente de la ciudad de Bogotá.

En resumen, significa que no existe correlación espacial entre las 19 estaciones de pluviometría del IDEAM, ubicadas en Bogotá, como también en los subgrupos de las 3 estaciones del nororiente, las 4 estaciones del centro, las 3 estaciones del sur y las 3 estaciones del occidente, todo esto cuando se estudia la variable precipitación, es decir la variable tiene un comportamiento aleatorio en cada uno de los grupos y subgrupos trabajados

El resultado anterior puede deberse a múltiples factores como: la existencia de micro climas en la ciudad, factores de contaminación, posibles fallas geológicas o el aumento de temperatura a nivel mundial, todo esto sería objeto de un estudio climatológico que se está fuera del alcance del presente trabajo.

9. CONCLUSIONES Y RECOMENDACIONES

9.1. CONCLUSIONES

- Al medir la correlación espacial mediante el índice I de Moran para datos funcionales, entre las 19 estaciones pluviométricas del IDEAM ubicadas en Bogotá, que tienen los registros más consistentes en un periodo de 1.232 días, comprendidos entre los años de 2.018 al 2.022 se concluye que no existe correlación espacial entre ellas.
- Al realizar la programación en R para poder medir el índice I de Moran para datos funcionales, definido en un artículo de estadística de universidades europeas y aplicado a las bases de datos de precipitación en la ciudad de Bogotá, se concluye que es completamente viable la aplicación de éste a las curvas de precipitación de la ciudad.
- Al depurar la base de datos de precipitaciones disponible en el IDEAM, que consta de 12 variables o columnas y aproximadamente 194.000.000 de registros o filas en una base de 19 variables y 1.232 registros se concluye que la imputación de la base de datos de precipitación en las 19 estaciones de pluviometría seleccionadas de la ciudad de Bogotá es satisfactoria.
- Se concluye que todos los resultados de la variación del índice I de Morán para datos funcionales, son similares cuando se cambia de método de medición realizado, y da igual resultado utilizando las distancias Euclidianas Sexagesimales y métricas, además de las distancias Manhattan, la Minkowski y la Chebysshev, Tanto en grupo general de 19 estaciones, como en los 5 subgrupos trabajados.

9.2. RECOMENDACIONES

- Se requiere una mayor cantidad de estaciones pluviométricas, (mínimo de 7), con fuente de energía de respaldo, especialmente en las localidades de Fontibón, Kennedy, Puente Aranda, Antonio Nariño, San Cristóbal, Santa Fe y Chapinero y de esta manera realizar registros más consistentes en toda la ciudad.
- Es necesario que los datos observados en las estaciones pluviométricas, tengan el mínimo posible de datos no disponibles o NA, para poder realizar estudios mucho más fiables y consistentes, pues de las 91 estaciones iniciales se tuvo que reducir a 19 el estudio; el 79 % de ellas no tenían datos consecutivos entre los años que se lleva el registro en el IDEAM.
- Es factible que las otras pruebas de autocorrelación espaciales, *(Geary's C, Join-count)* puedan modificarse para ser aplicadas a datos funcionales, lo que podría abrir otros campos de interés para realizar trabajos de grado. No porque sean más adecuadas o no, solo por el avance estadístico aplicado a la cada vez mayor cantidad de datos funcionales que se requiere en el desarrollo tecnológico de la información.

Referencias

Anselin, L. (1988), *Econometría espacial: métodos y modelos. Kluwer Academic, Dordrecht.*

Armstrong, A. (((1998)), *Basic Linear Geostatistics. Springer. New York.*

Balzanella A, s.a. gattone, t. d. b. e. r. r. s. (2017), *statistics and data science: new challenges, new generations 28–30 june 2017 florence (italy) proceedings of the conference of the italian statistical society edited by alessandra petrucci rosanna verde firenze university press 2017 sis 2017. statistics and data science: new challenges, new.*

Berrocal, V.J., G. A. E. H. D. ((2010)), *A spatio-temporal downscaler for outputs from numerical models. JABES. 15(2): 176-197. .*

Cameletti, M., L. F. S. D. R. H. ((2013).), *A spatio-temporalmodeling of particulate matter concentration. Advances in Statistical Analysis, 97(2): 109-131.*

CK-12 Foundation (n.d.), 'CK12-foundation', `https://flexbooks.ck12.org/cbook/ck-12-conceptos-de-ciencias-de-la-tierra-grados-6-8-en-espanol/section/8.5/primary/lesson/precipitaciones/`. Accessed: 2025-6-9.

Cohen, J. ((1997).), *Much about nothing. Lecture presented at the annual meeting of the American Psychological Association, Chicago.*

Cressi, n. y. w. c. (2011), *statistics spatio-temporal data: ed. wiley.*

Cuadras, C. ((1989)), *Distancias Estadísticas. ESTADISTICA ESPAÑOLA Vol. 30, Núm. 1 19, 1989, págs. 295 a 378.*

De Boor, C. (2001), *Una guía práctica para splines (Ciencias matemáticas aplicadas). Springer-Verlag.*

Gittleman, JL, y. K. M. ((1990).), *Adaptación: estadísticas y un modelo nulo para estimar los efectos filogenéticos Zoología sistemática.*

Grenander, U. . (1950), *Stochastic processes and stadistical inference. Arkiv för matematik.*

Guevara, r. (2015), *notas de clase. universidad nacional de colombia.*

Horvath lajos, k. ((2012).), *inference for functional data with applications. new york: springer.*

IDEAM (2018), *ideam.gov.co.Recuperado de https://datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Precipitaci-n/s54a-sgyg/about$_d$ata.*

Instagram (n.d.), `https://www.instagram.com/p/CeynUTUu93O/`. Accessed: 2025-6-9.

Jerrett, M., A. M. K. P. ((2005)), *A review and evaluation of intraurban air pollution exposure models. Journal of Exposure Science and Environmental Epidemiology, 15(2): 185-204 .*

Kokoszka, P y Reimherr, M. (2017), *Introduction to Functional Data Analysis.*, New York: Taylor y Francis.

Kopczewska, k. (2020), *estadística espacial aplicada y econometría análisis de datos en r: ed. routledge.*

Lawson, A. ((2013).), *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology, CRC Press.*

Martori, J. (1999), *Modelización econométrica de la densidad de población urbana: la aproximación clásica. Evidencia empírica para 40 ciudades catalanas.*

Miller, A. A. (1996), *Climatología, Geografía., Barcelona, Omega.*

MinutoUno (2024), 'Ola de frío: qué diferencia hay entre nieve, aguanieve, graupel y lluvia helada', `https://www.minutouno.com/sociedad/nieve/ola-frio-que-diferencia-hay-agua-graupel-y-lluvia-helada-n6027543`. Accessed: 2025-6-9.

Ramsay .J. O, S. B. (1989), *Functional Data Analysis Springer.*

Rao, C. (1958), *Bengal anthropometric survey 1945: A statistical study. Sankhya,.*

Raya, a. (2007), *introducción a los espacios de hilbert: ed. @becedario.*

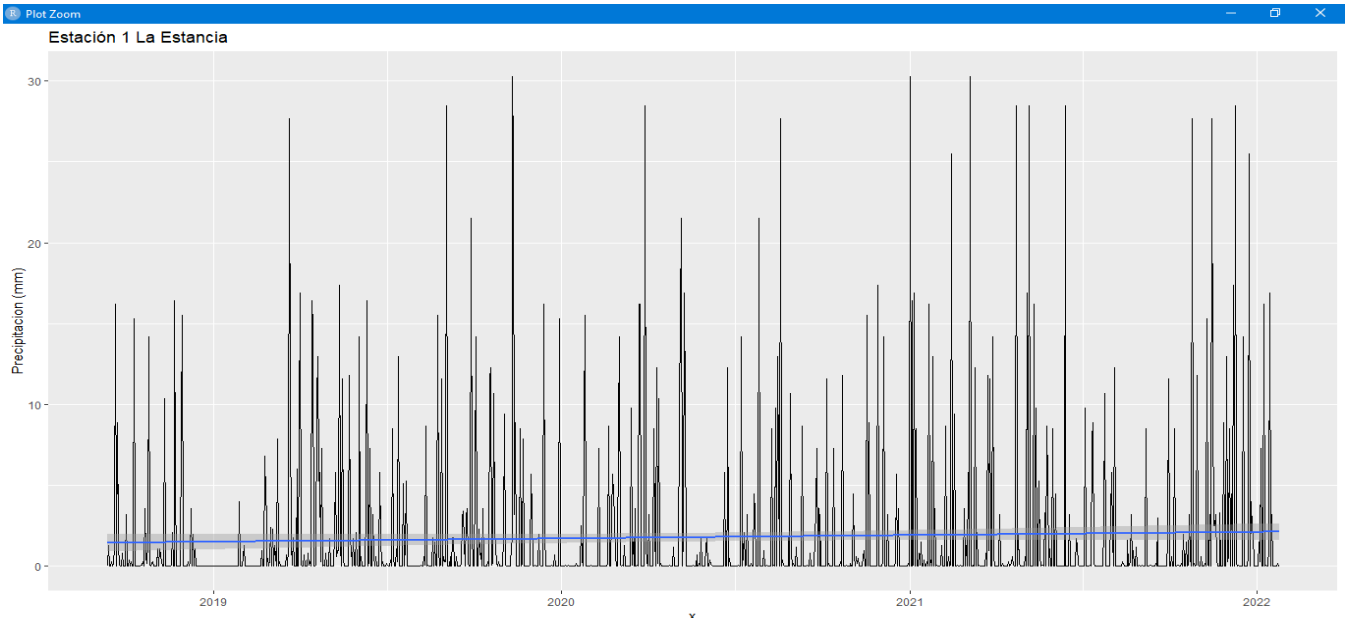
Schabenberger, O.y Pierce, F. (2001), *Contemporary Statistical Models for the Plant and Soil Sciences. Ed, Crc Press.*

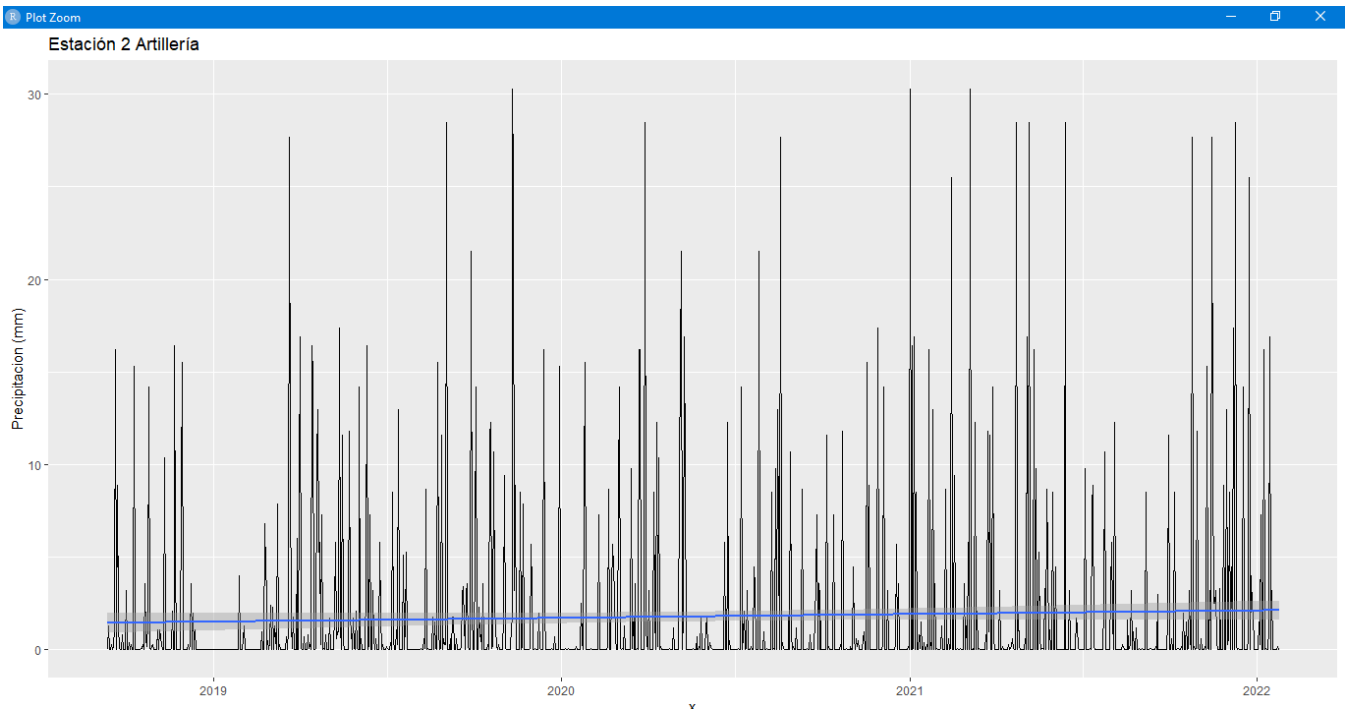
Vargas (2011), *Análisis de la distribución e interpolación espacial de las lluvias en bogotá, colombia: Ed. Universidad Javeriana.*

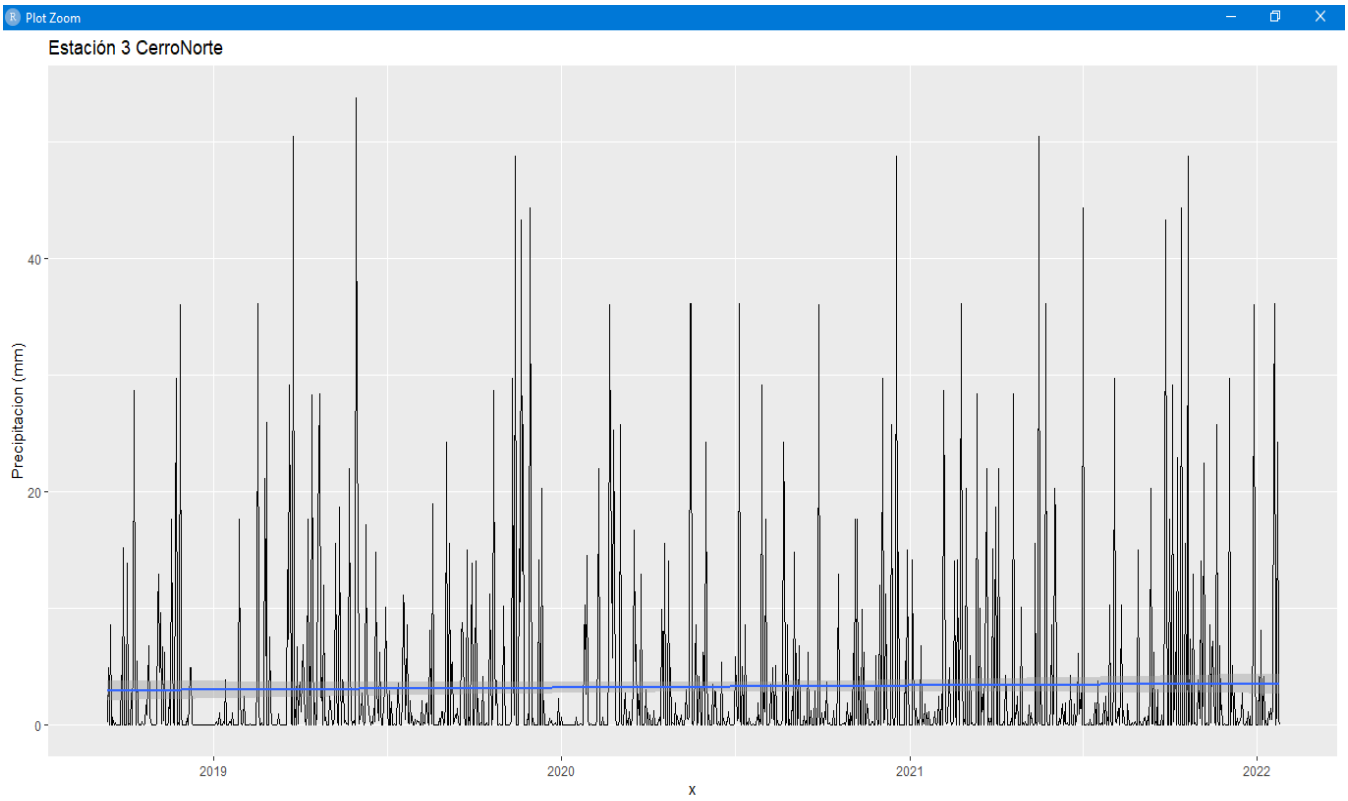
Ward, R. C., . R. M. (((2000)), *Principles of Hydrology (4th ed.). McGraw-Hill.*

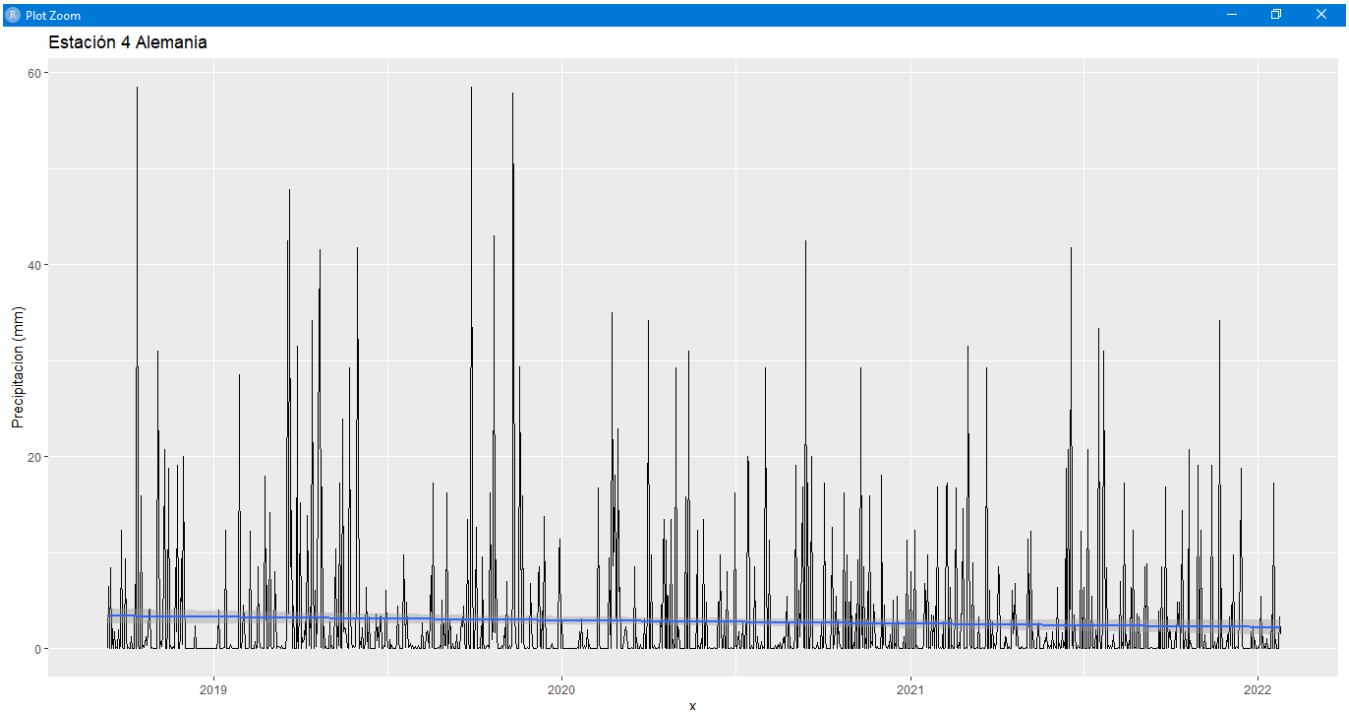
Xavier, G. C. (2010), *Spatial Statistics and Modeling.*, New York: Taylor y Francis.

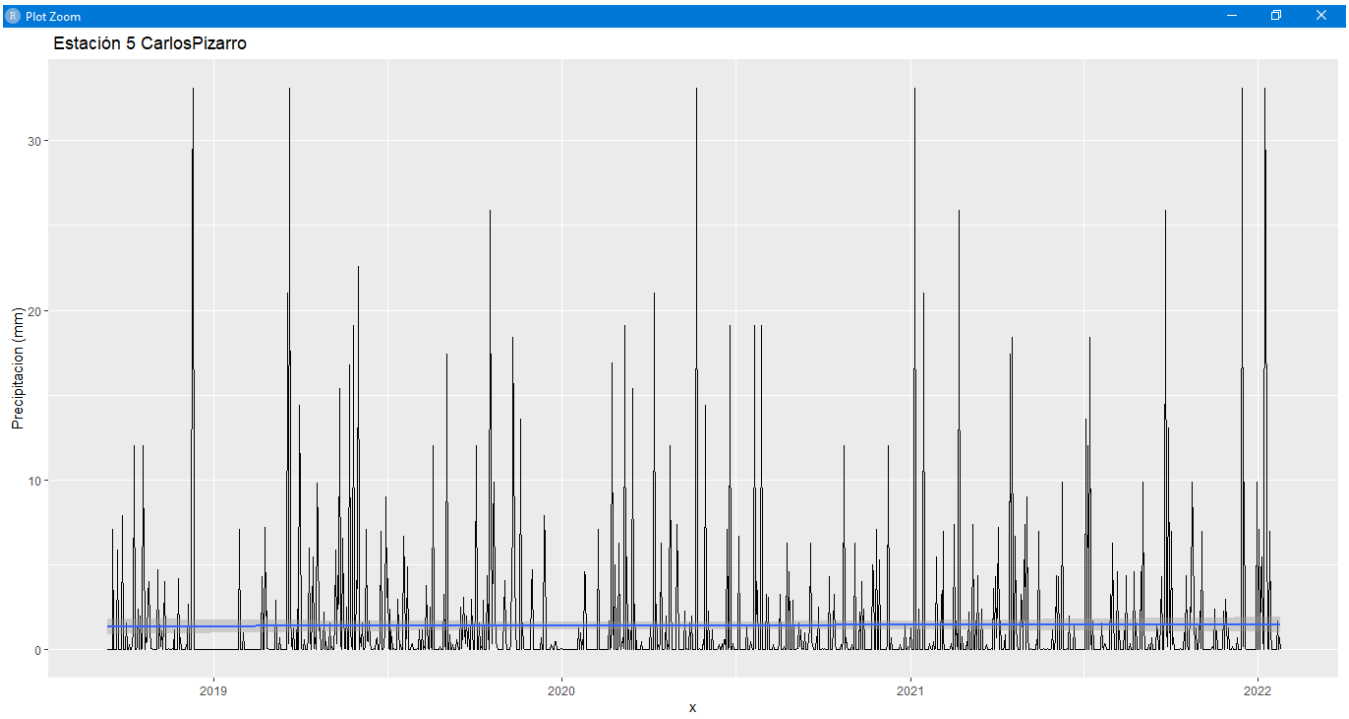
Anexo 1: Gráficas de las precipitaciones en las 19 estaciones pluviométricas seleccionadas y de Bogotá, como datos discretos y con la línea de la media en azul.

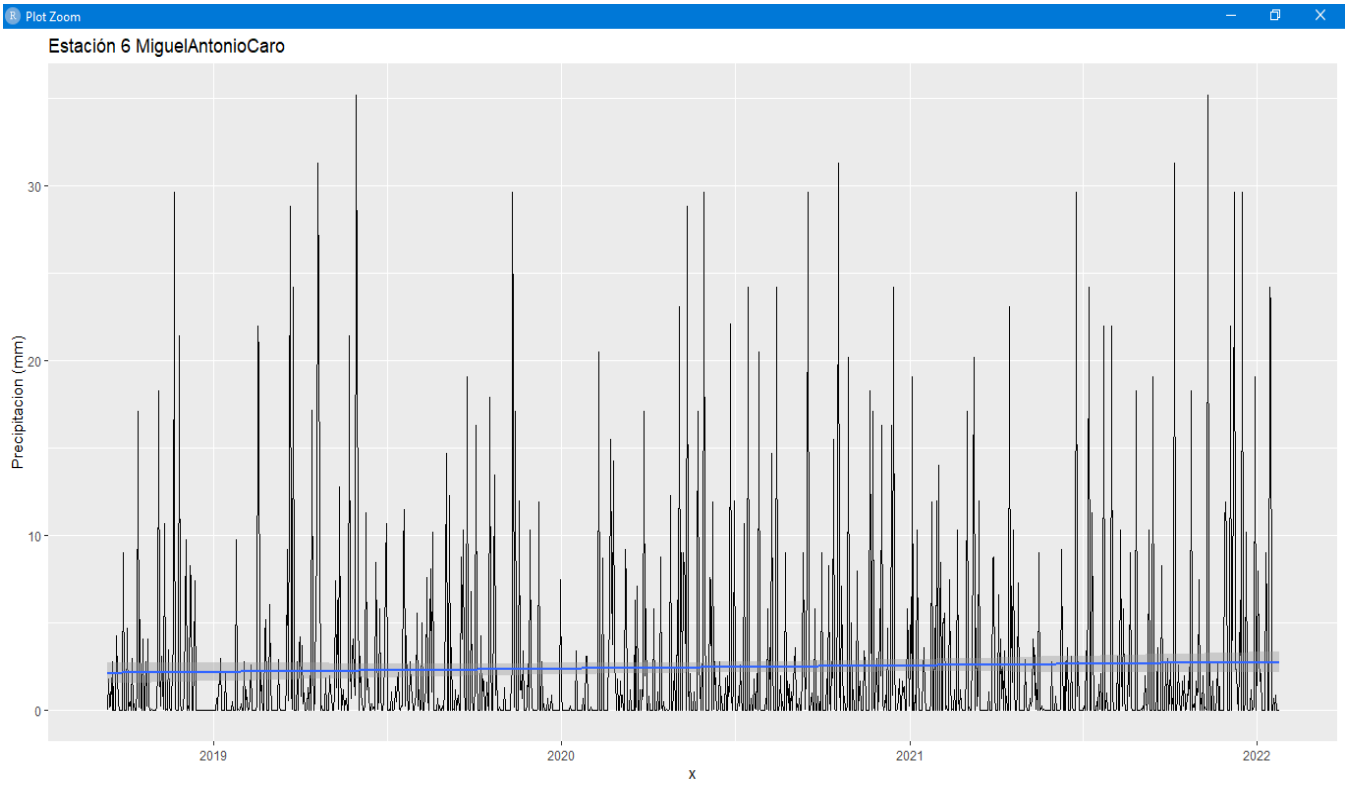


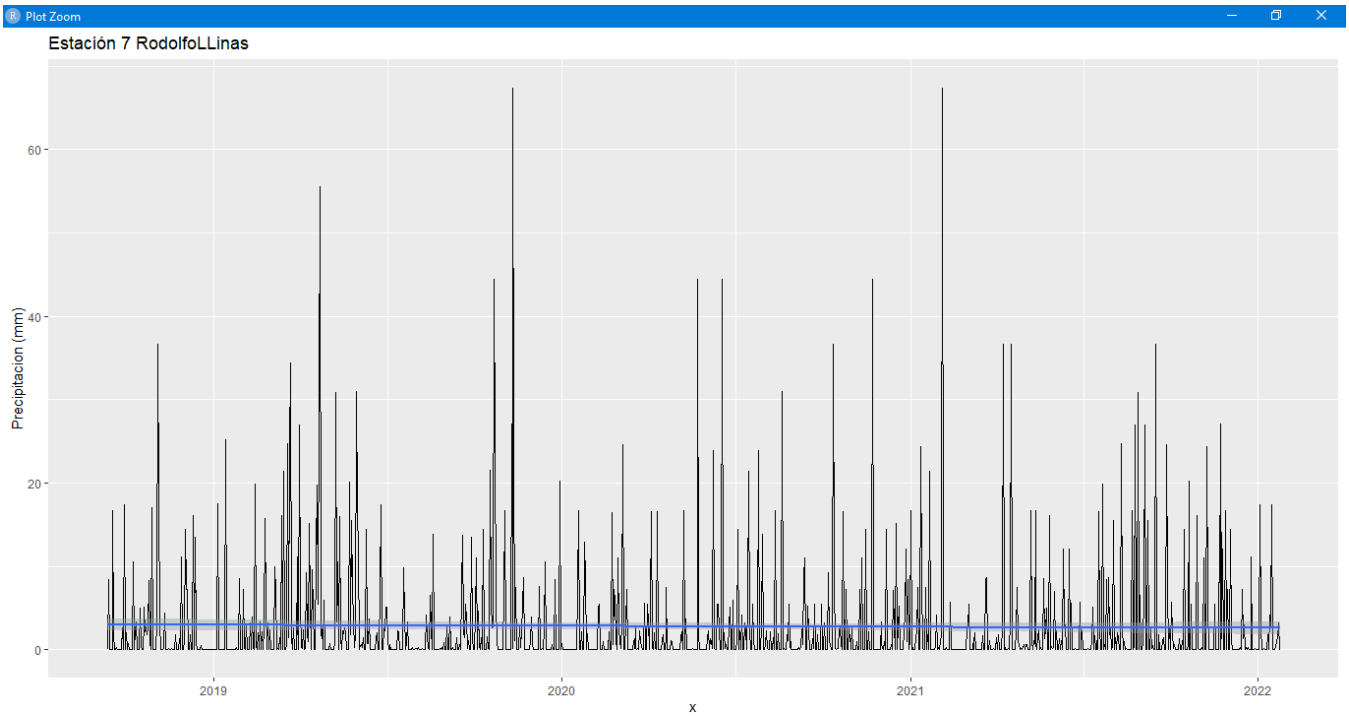


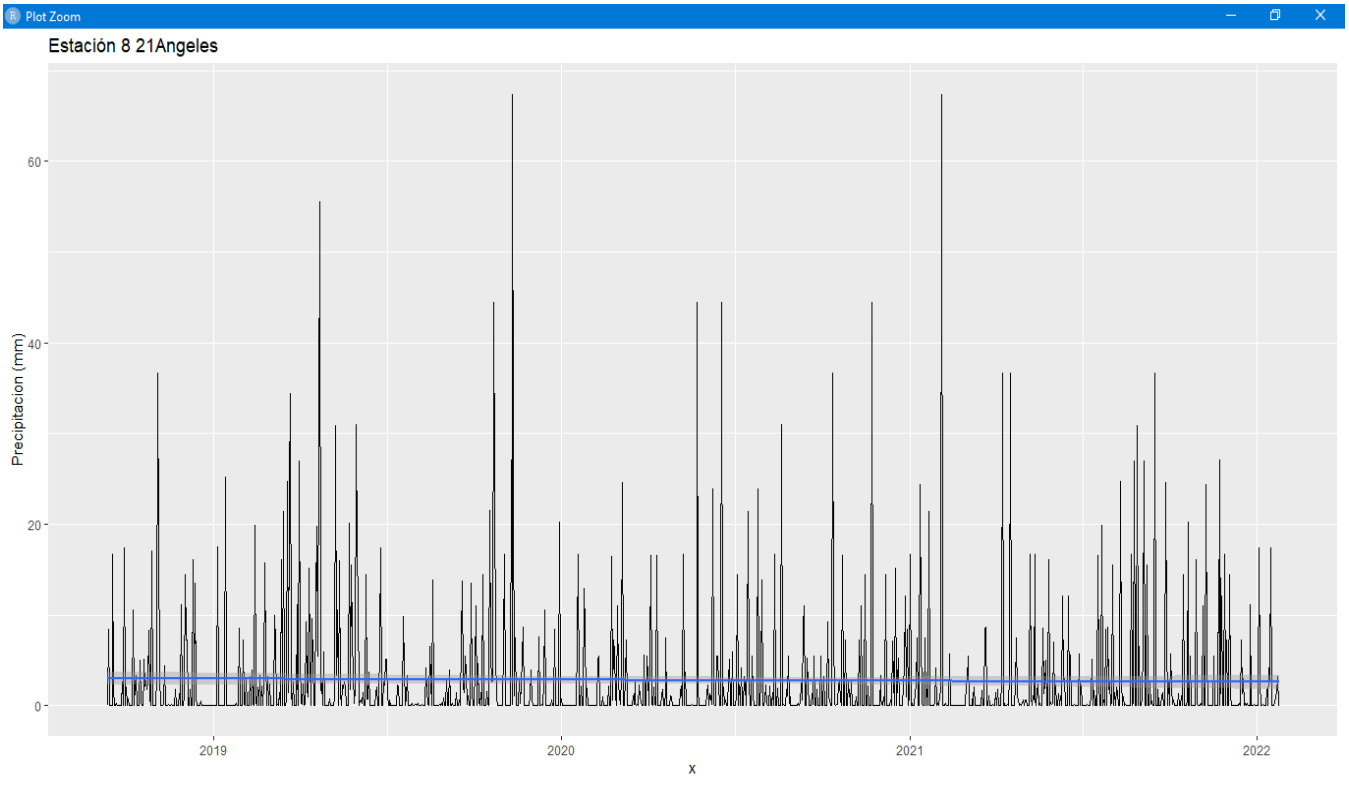

Plot Zoom
Estación 3 CerroNorte
Precipitacion (mm)
0
20
40
2019
2020
2021
2022
x

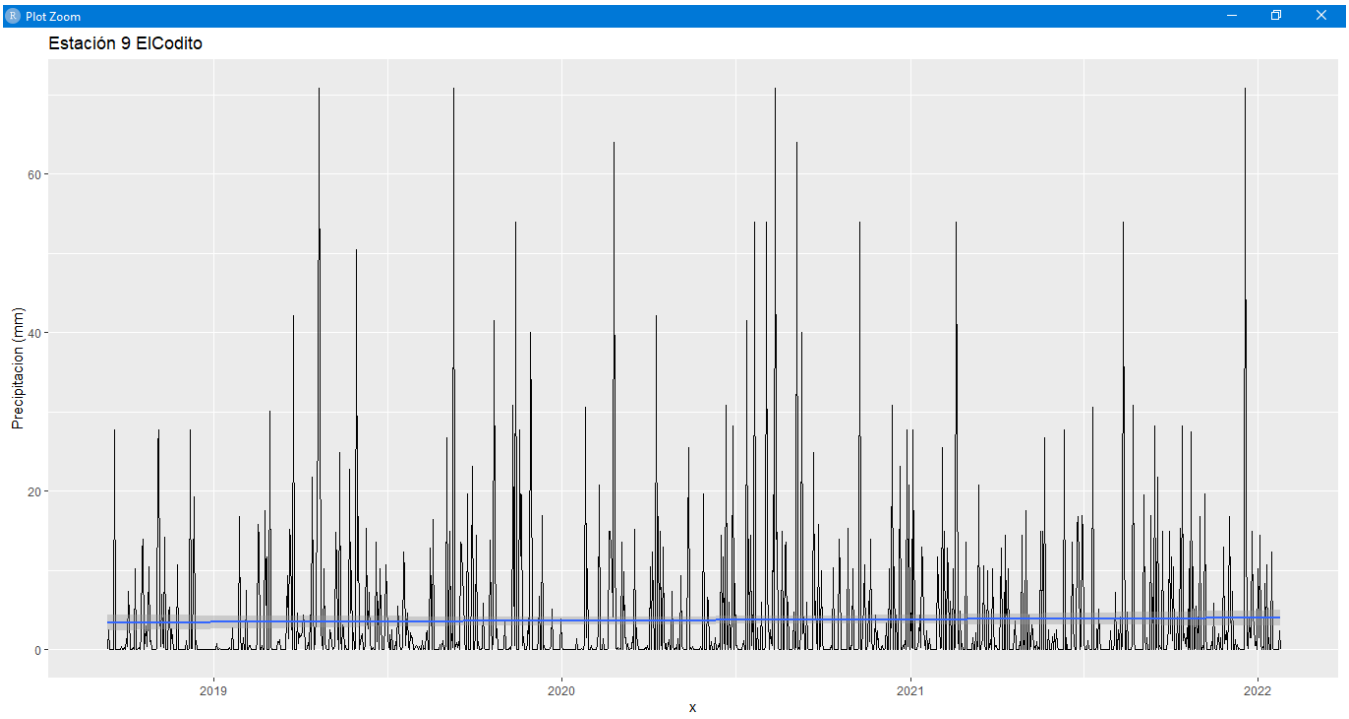

Plot Zoom
Estación 4 Alemania
Precipitacion (mm)
0
20
40
60
2019
2020
2021
2022
x

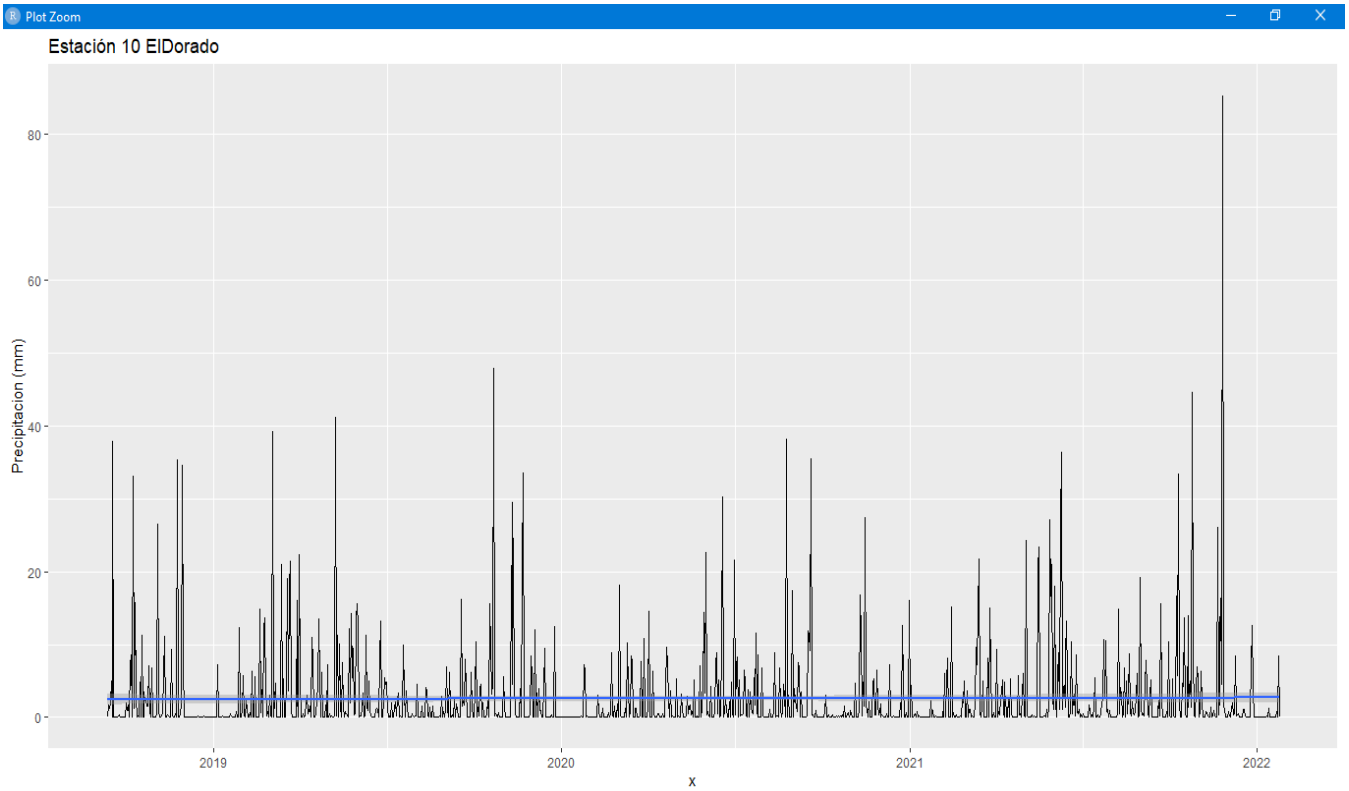

Plot Zoom
Estación 5 CarlosPizarro
Precipitacion (mm)
30
20
10
0
2019
2020
2021
2022
x

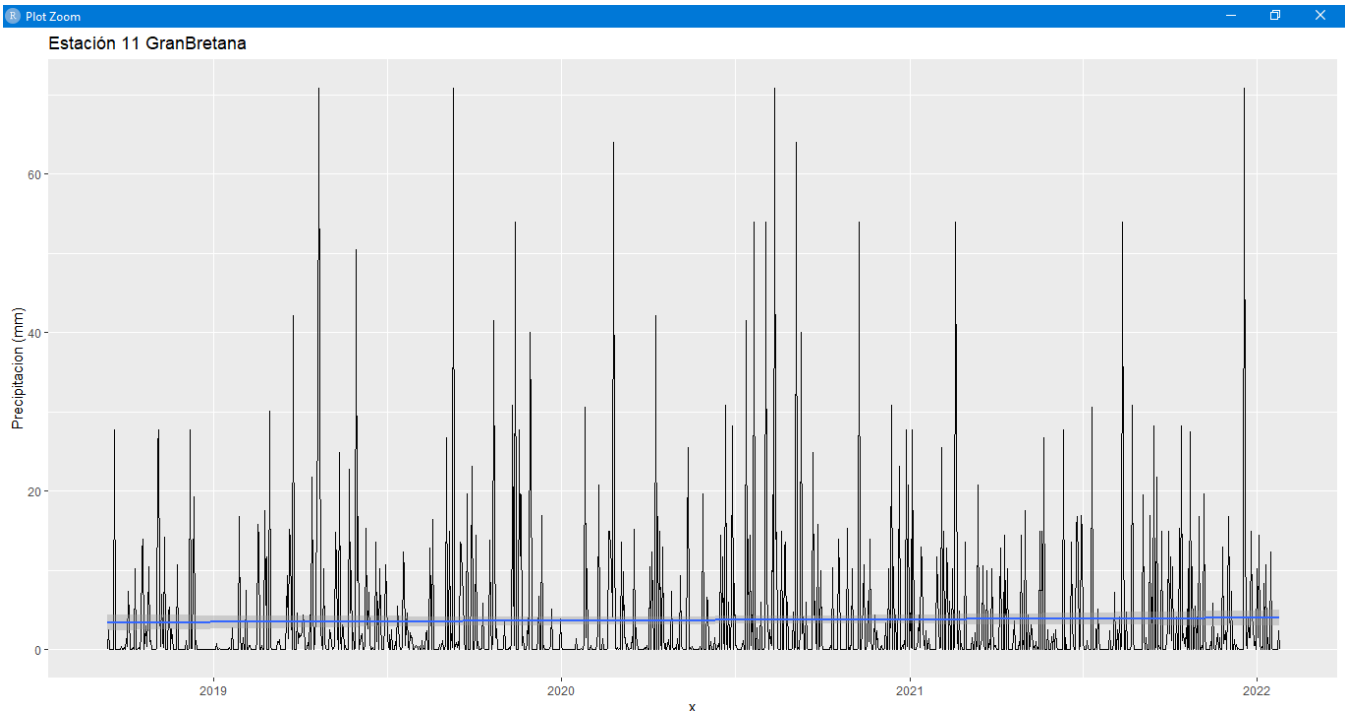

Plot Zoom
Estación 6 MiguelAntonioCaro
Precipitacion (mm)
30
20
10
0
2019
2020
2021
2022
x

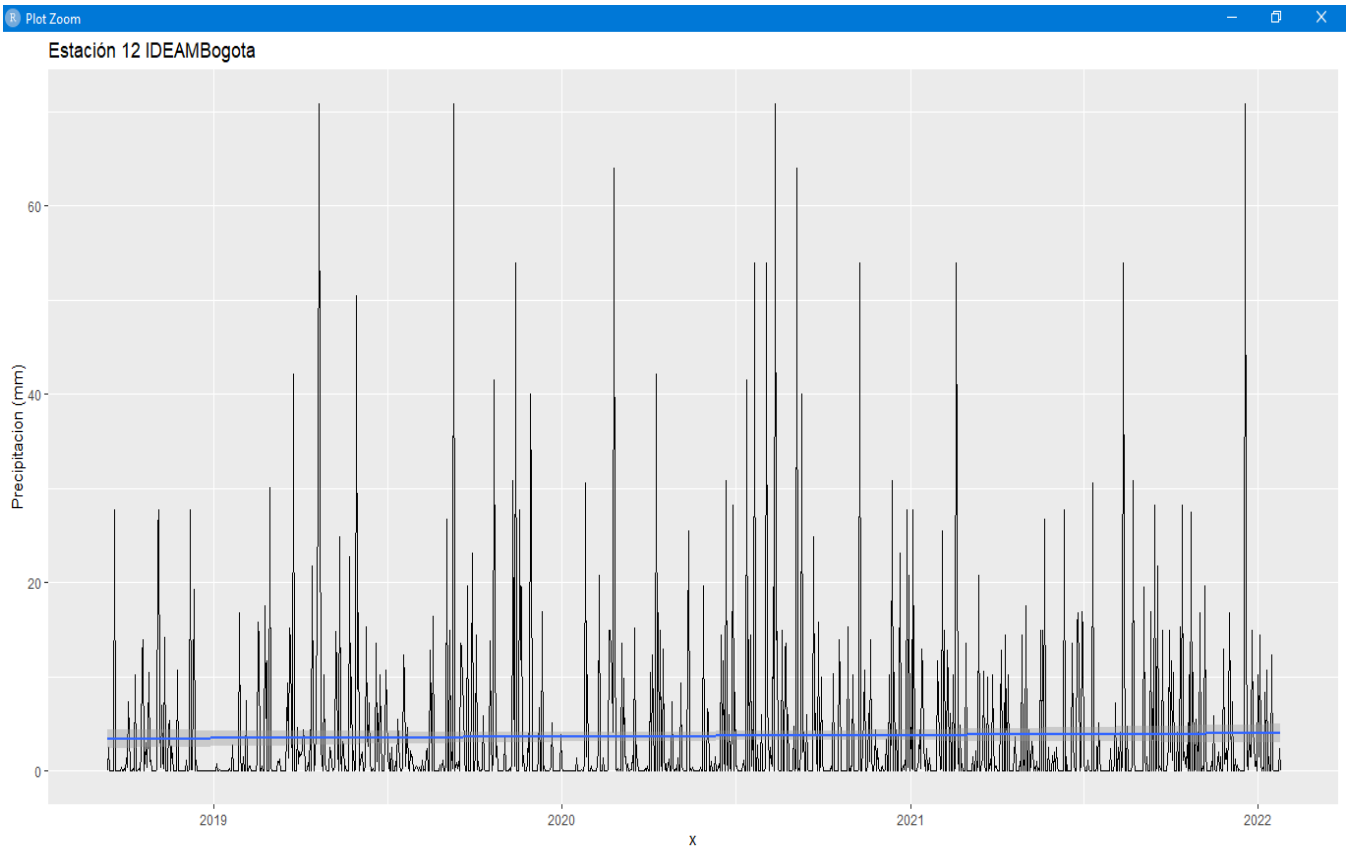

Plot Zoom
Estación 7 RodolfoLLinas
Precipitacion (mm)
0
20
40
60
2019
2020
2021
2022
x

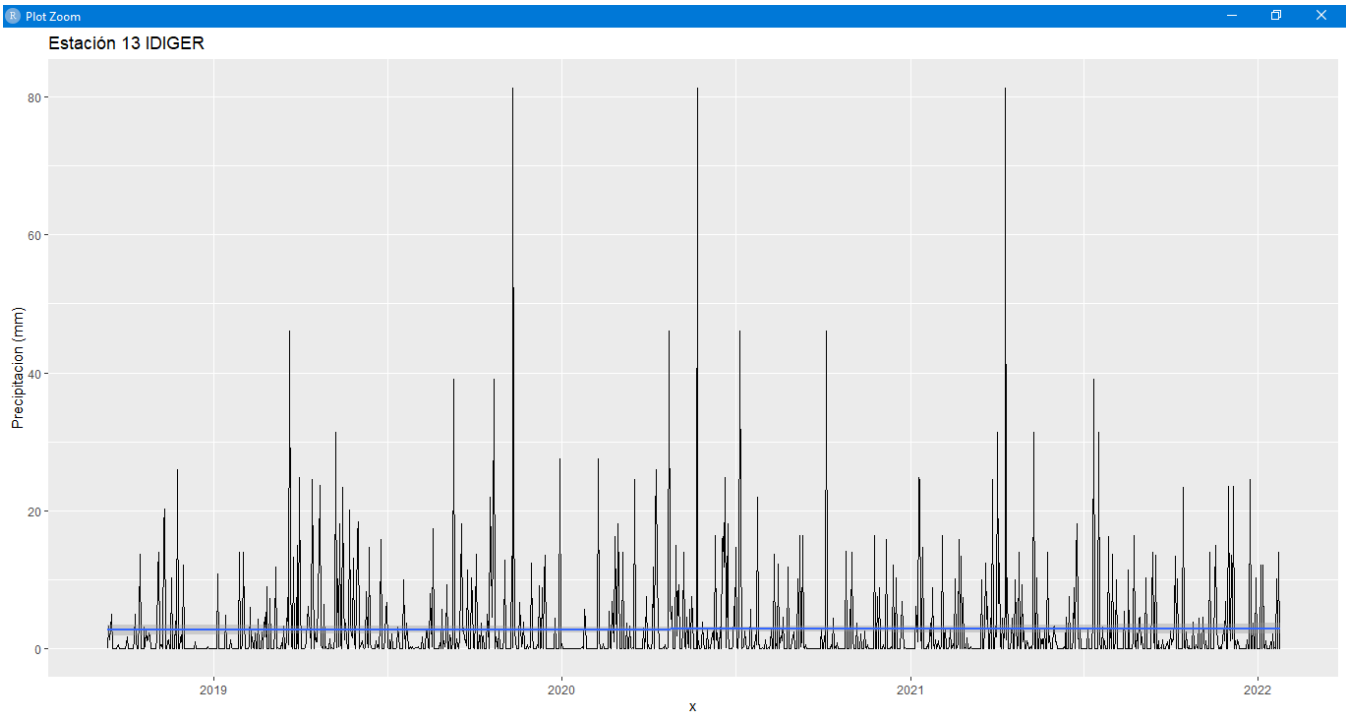

Plot Zoom
Estación 8 21Angeles
Precipitacion (mm)
0
20
40
60
2019
2020
2021
2022
x

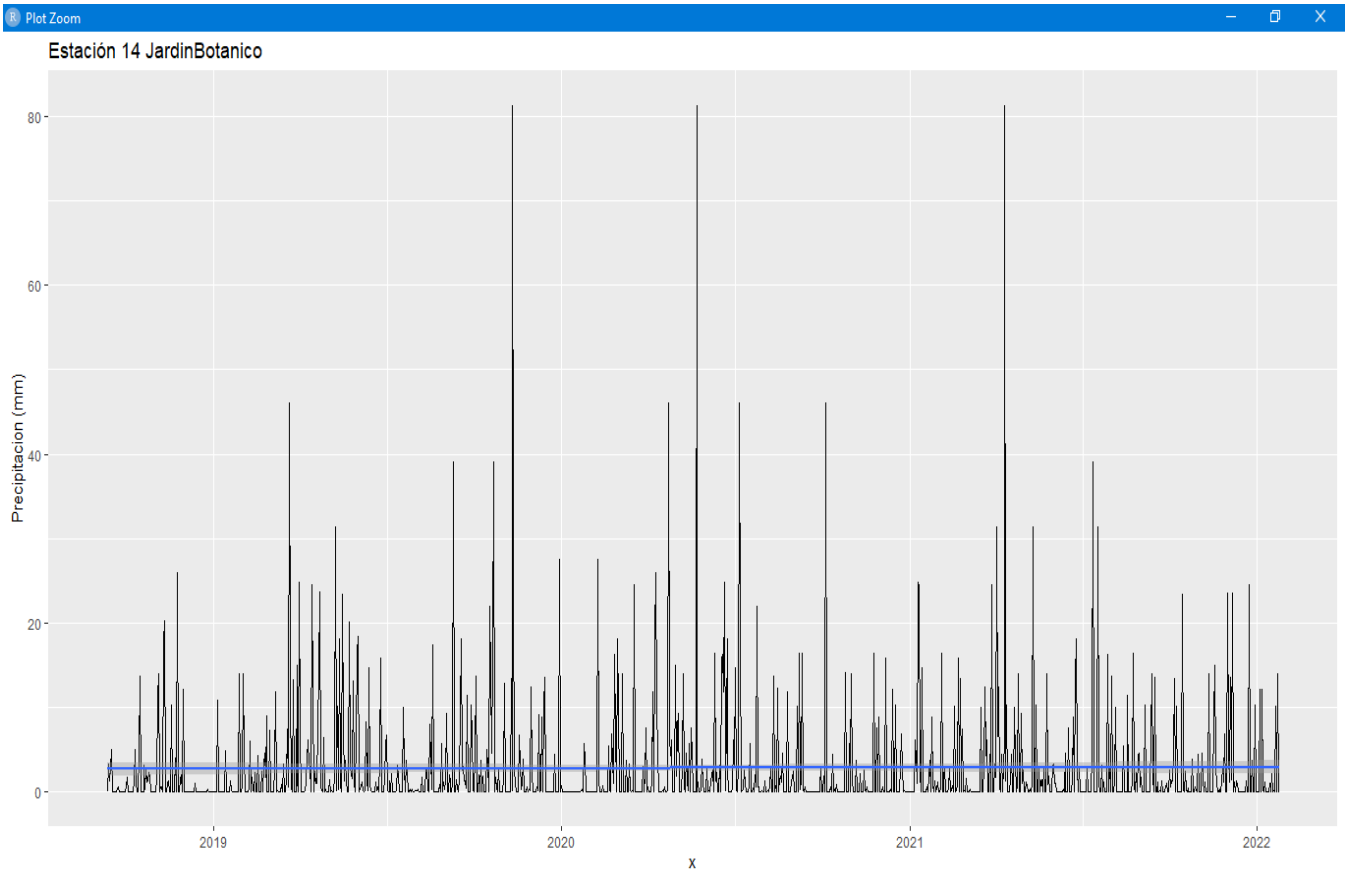

Plot Zoom
Estación 9 ElCodito
Precipitacion (mm)
0
20
40
60
2019
2020
2021
2022
x

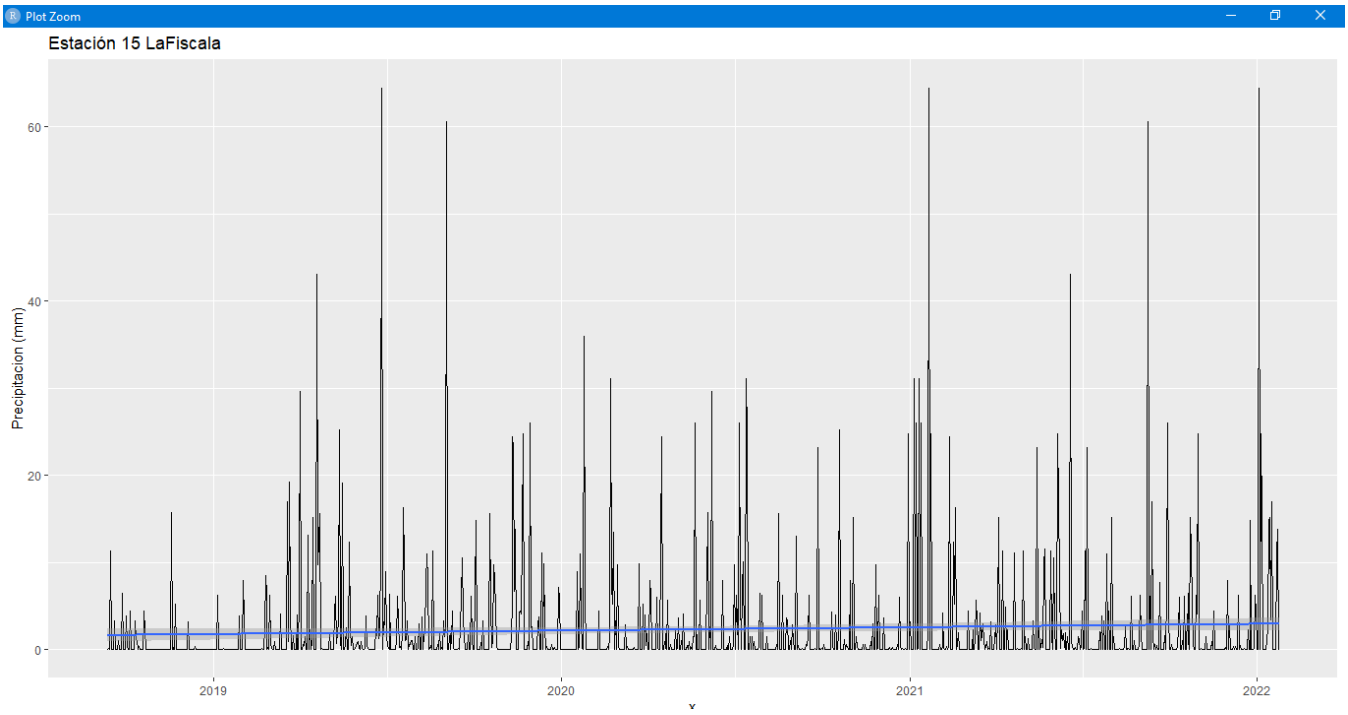

Plot Zoom
Estación 10 ElDorado
Precipitacion (mm)
0
20
40
60
80
2019
2020
2021
2022
x

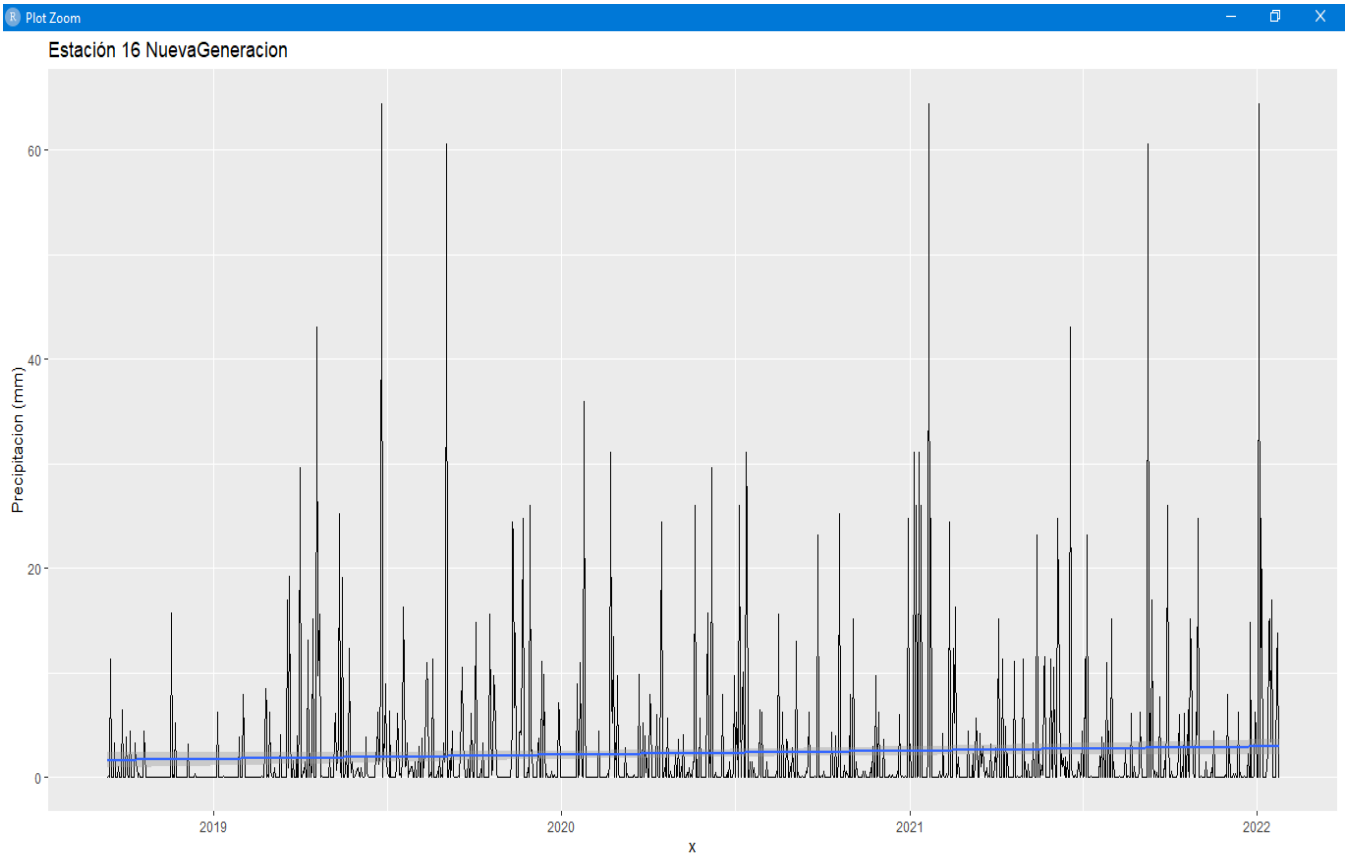

Plot Zoom
Estación 11 GranBretana
Precipitacion (mm)
0
20
40
60
2019
2020
2021
2022
x

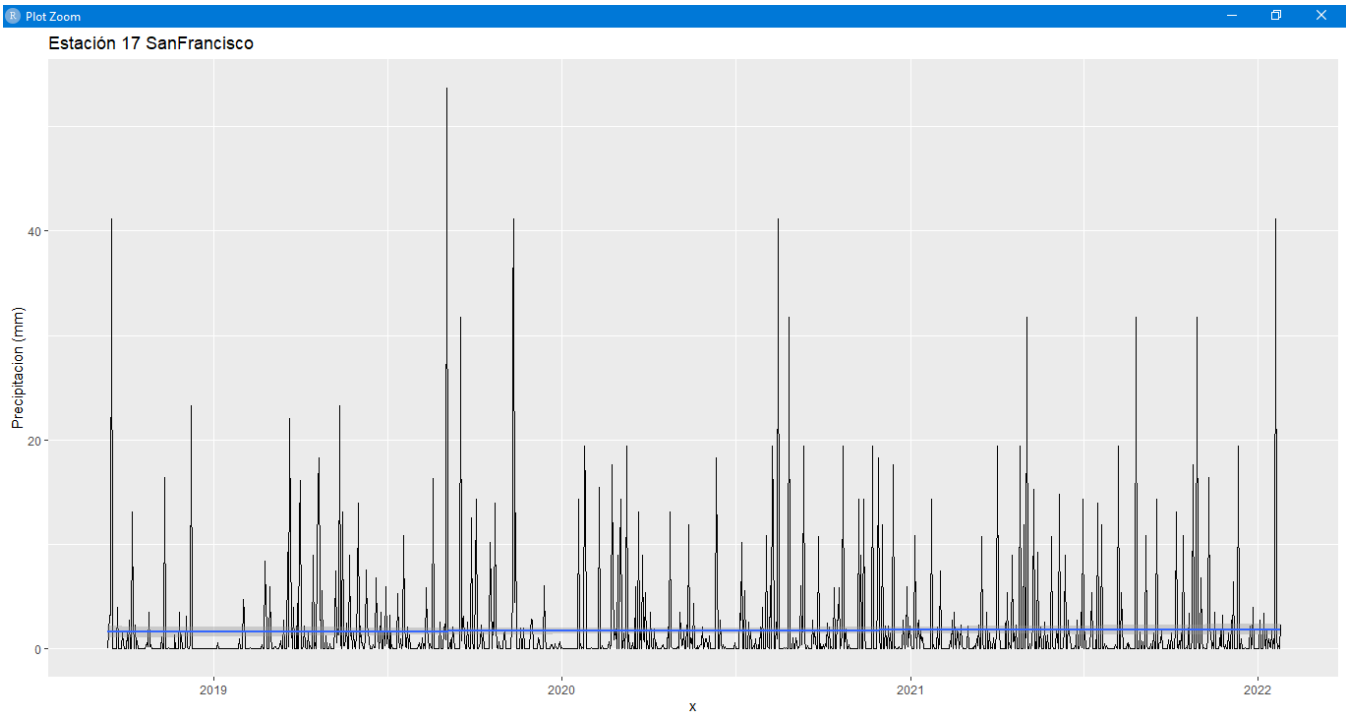

Plot Zoom
Estación 12 IDEAMBogota
Precipitacion (mm)
0
20
40
60
2019
2020
2021
2022
x

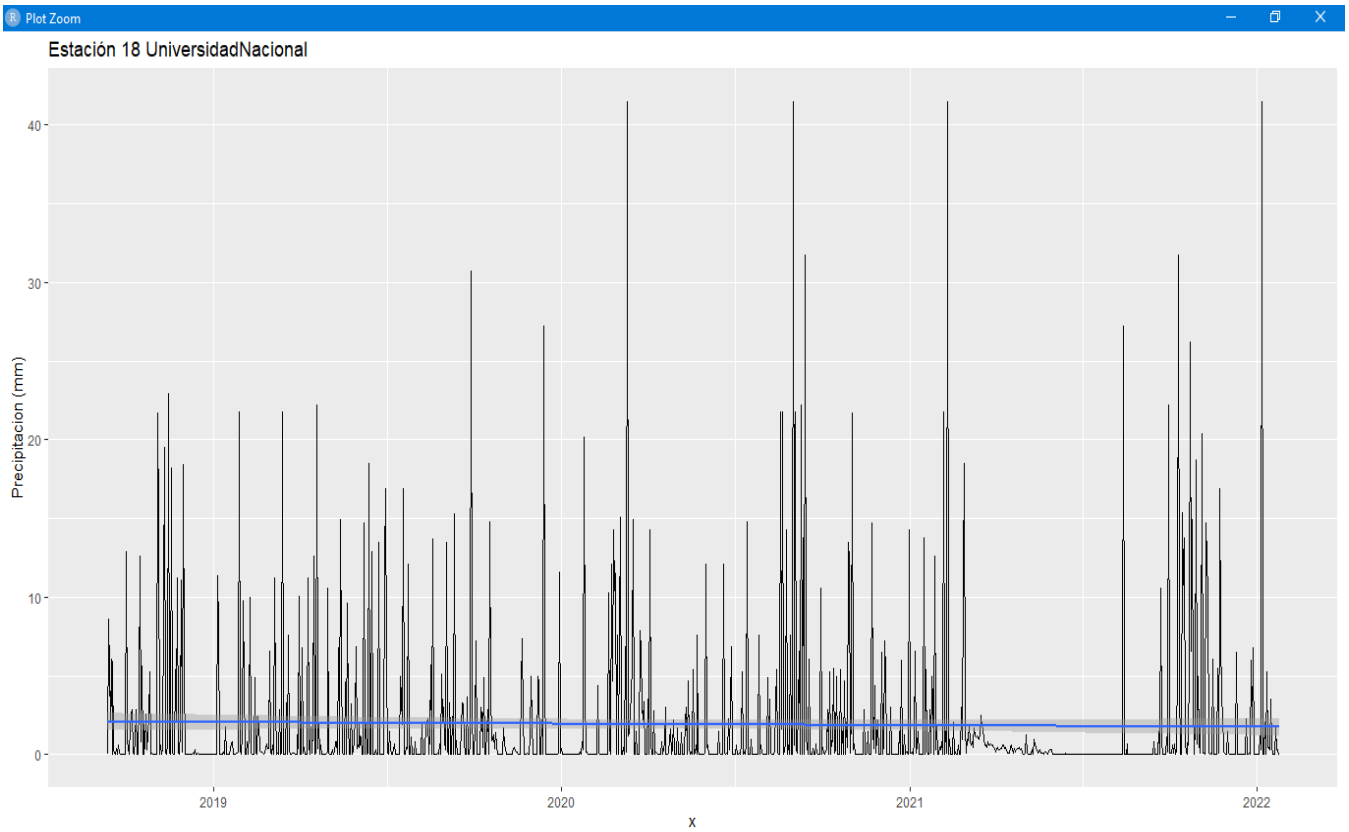

Plot Zoom
Estación 13 IDIGER
Precipitacion (mm)
0
20
40
60
80
2019
2020
2021
2022
x

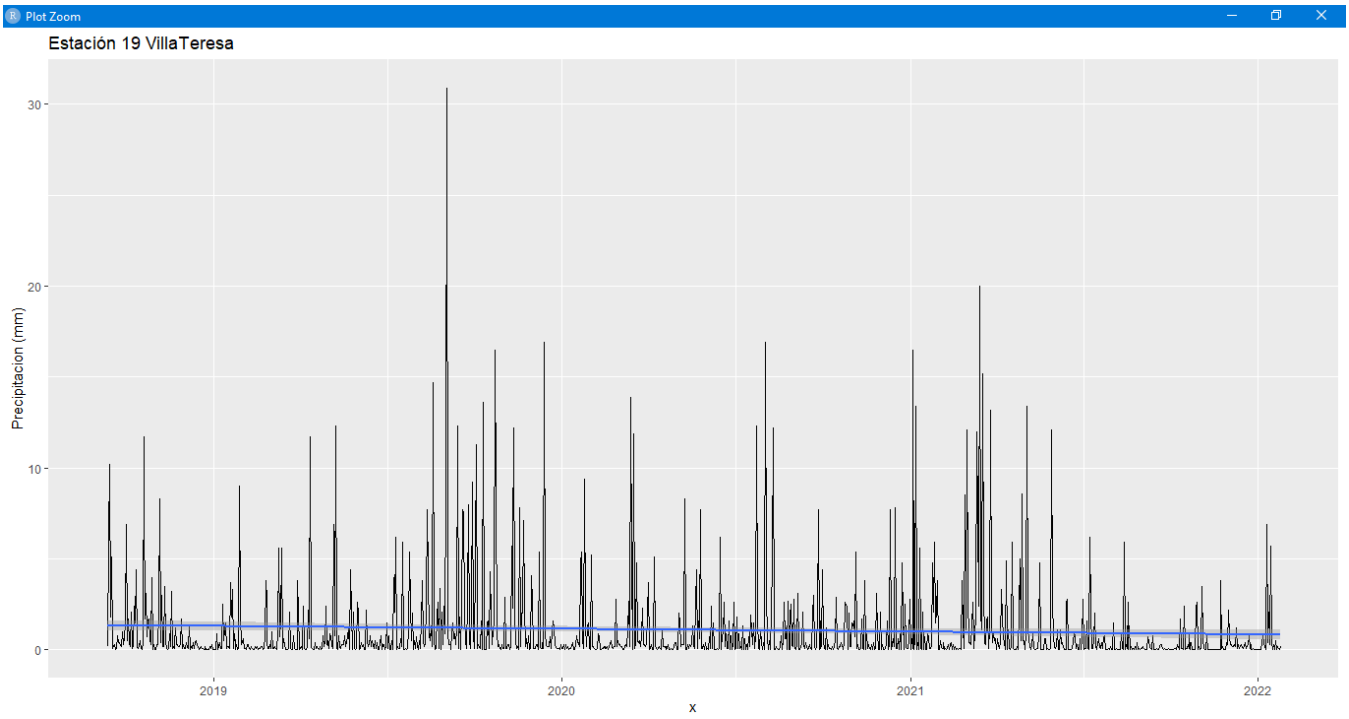

Plot Zoom
Estación 14 JardinBotanico
Precipitacion (mm)
0
20
40
60
80
2019
2020
2021
2022
x

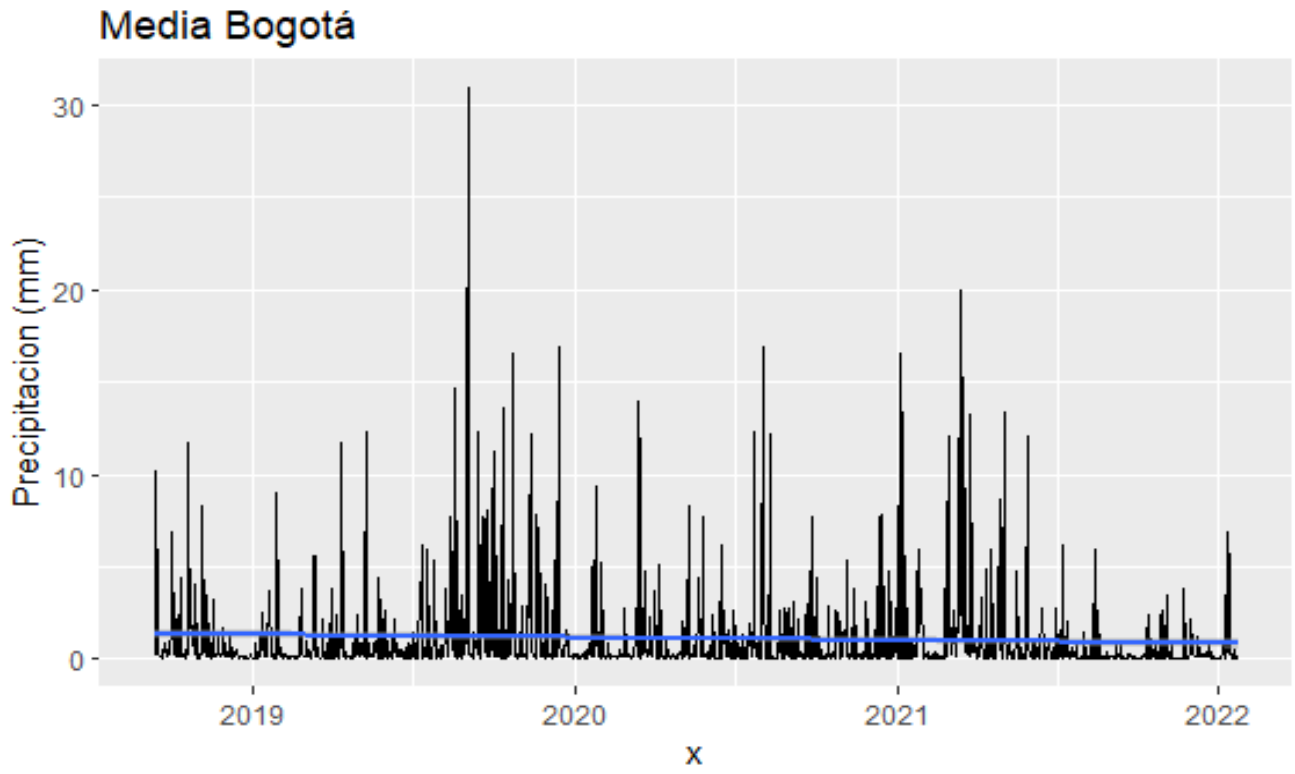

Plot Zoom
Estación 15 LaFiscala
Precipitacion (mm)
0
20
40
60
2019
2020
2021
2022
x


Plot Zoom
Estación 16 NuevaGeneracion
Precipitacion (mm)
0
20
40
60
2019
2020
2021
2022
x


Plot Zoom
Estación 17 SanFrancisco
Precipitacion (mm)
0
20
40
2019
2020
2021
2022
x

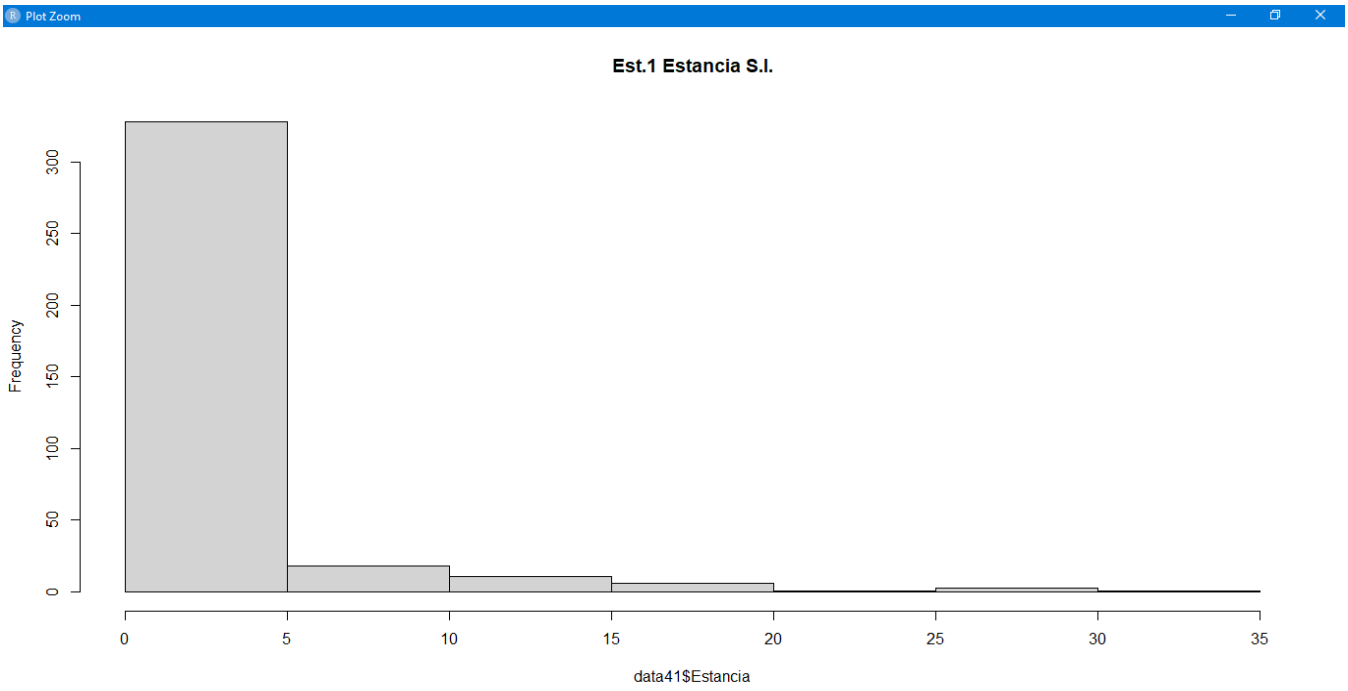

Plot Zoom
Estación 18 UniversidadNacional
Precipitacion (mm)
0
10
20
30
40
2019
2020
2021
2022
x

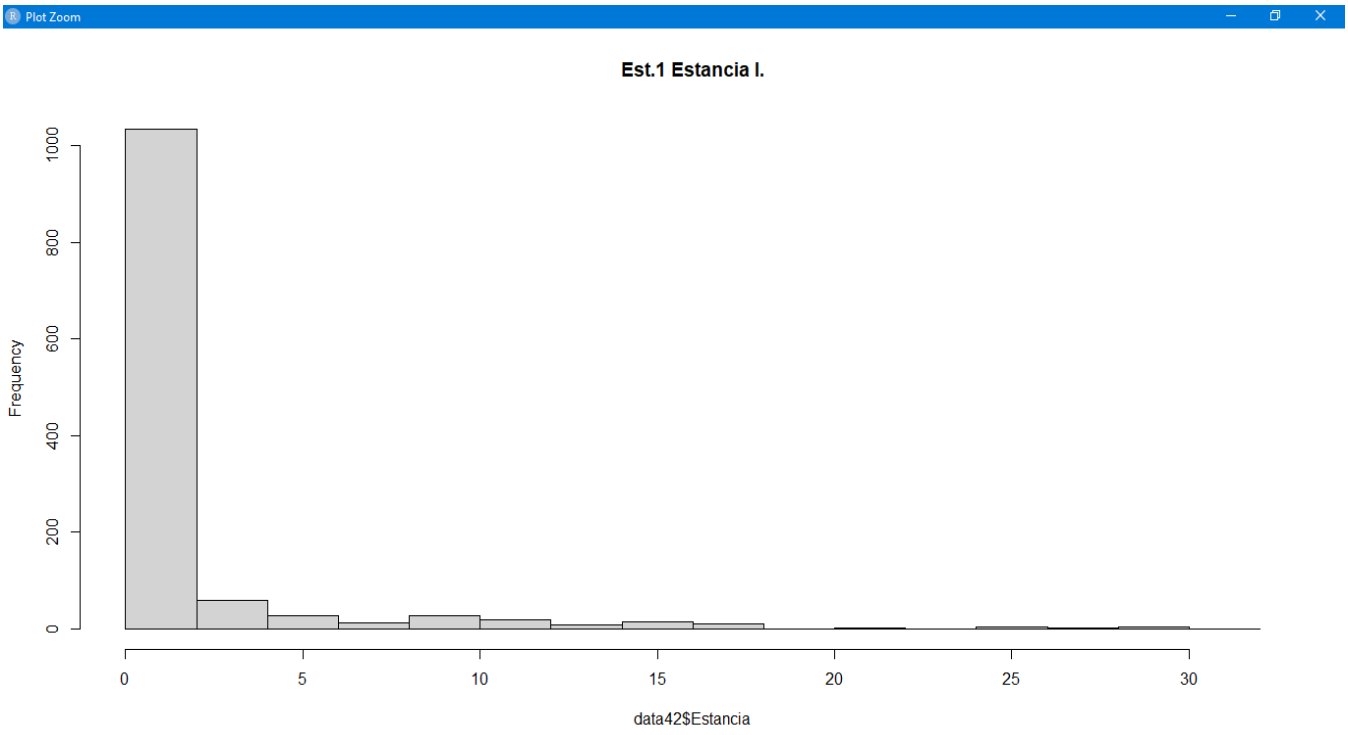

Plot Zoom
Estación 19 VillaTeresa
Precipitacion (mm)
30
20
10
0
2019
2020
2021
2022
x


Media Bogotá
Precipitacion (mm)
30
20
10
0
2019
2020
2021
2022
x

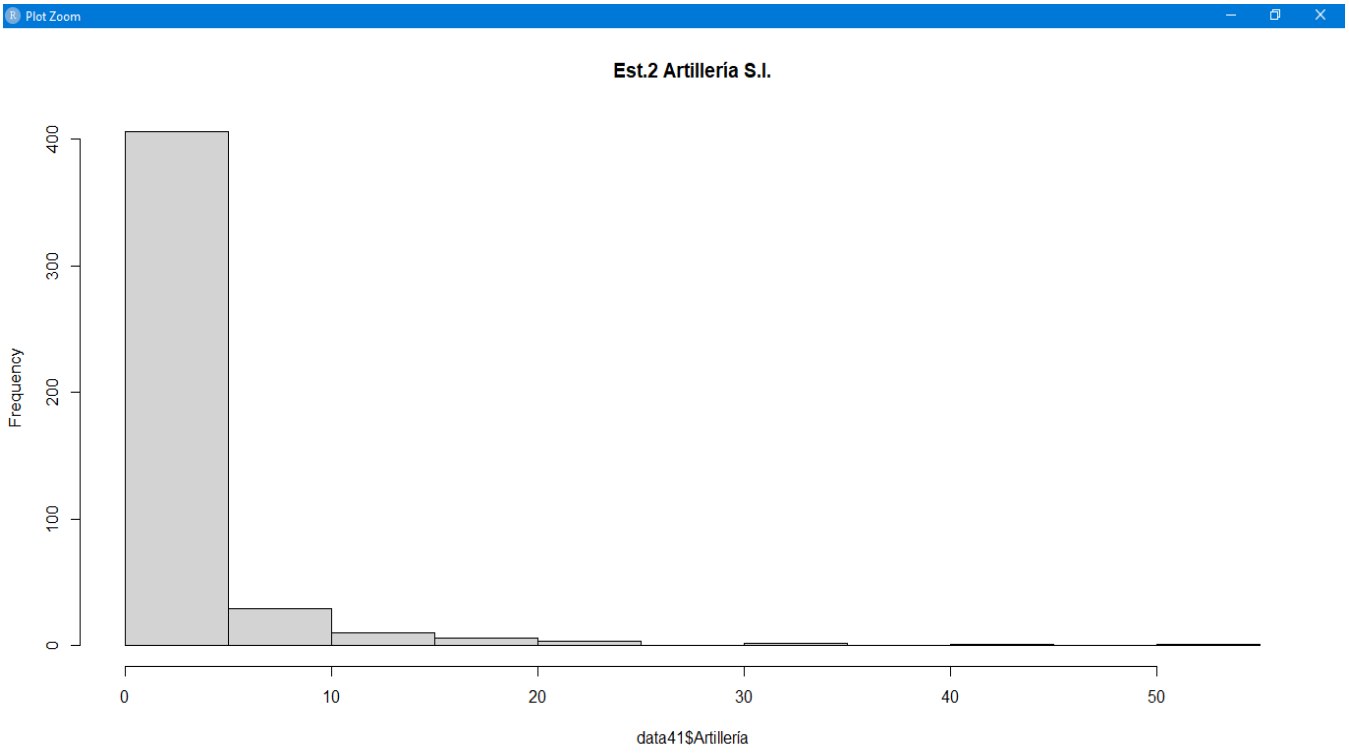
Anexo 2: Histogramas comparativos de las precipitaciones en las 19 estaciones pluviométricas seleccionadas antes (S.I.) y después de ser imputadas (I.)

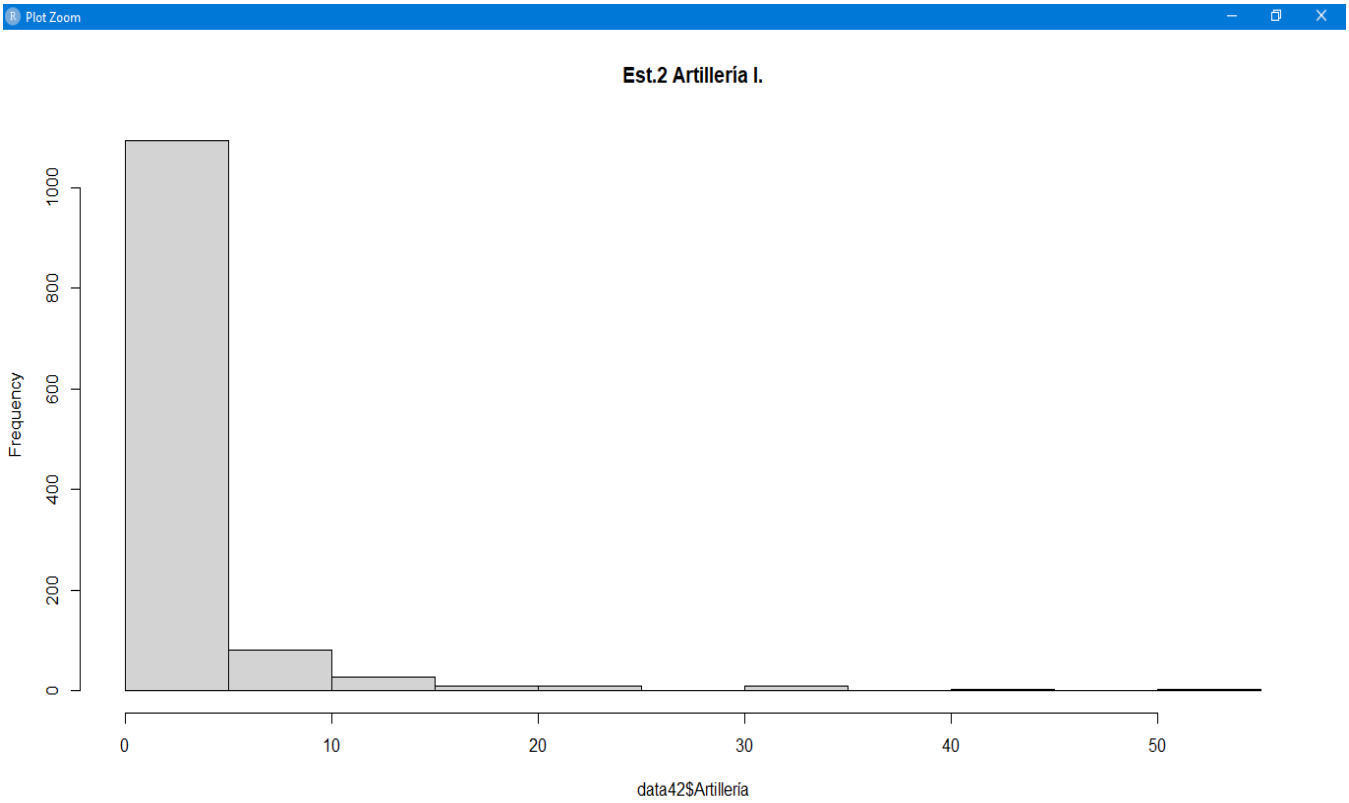
N.1 Estación Estancia



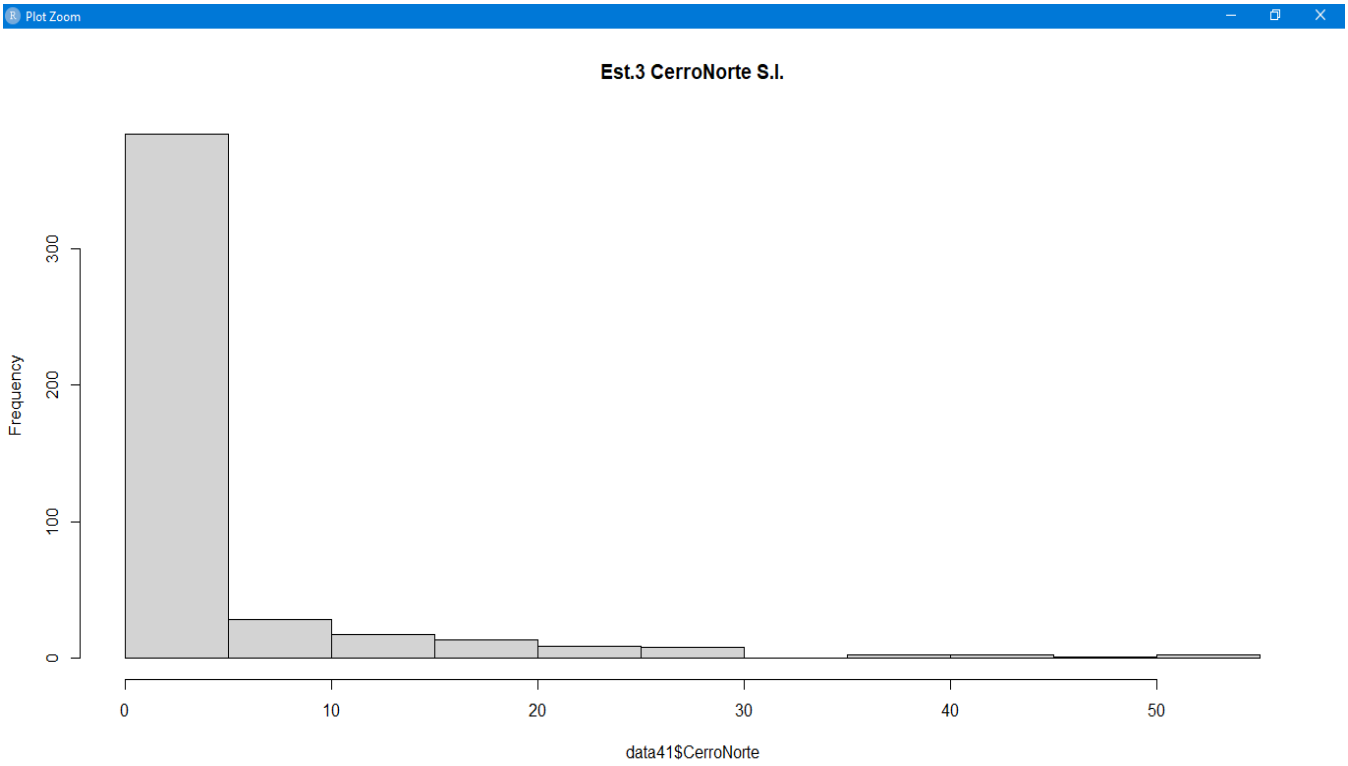


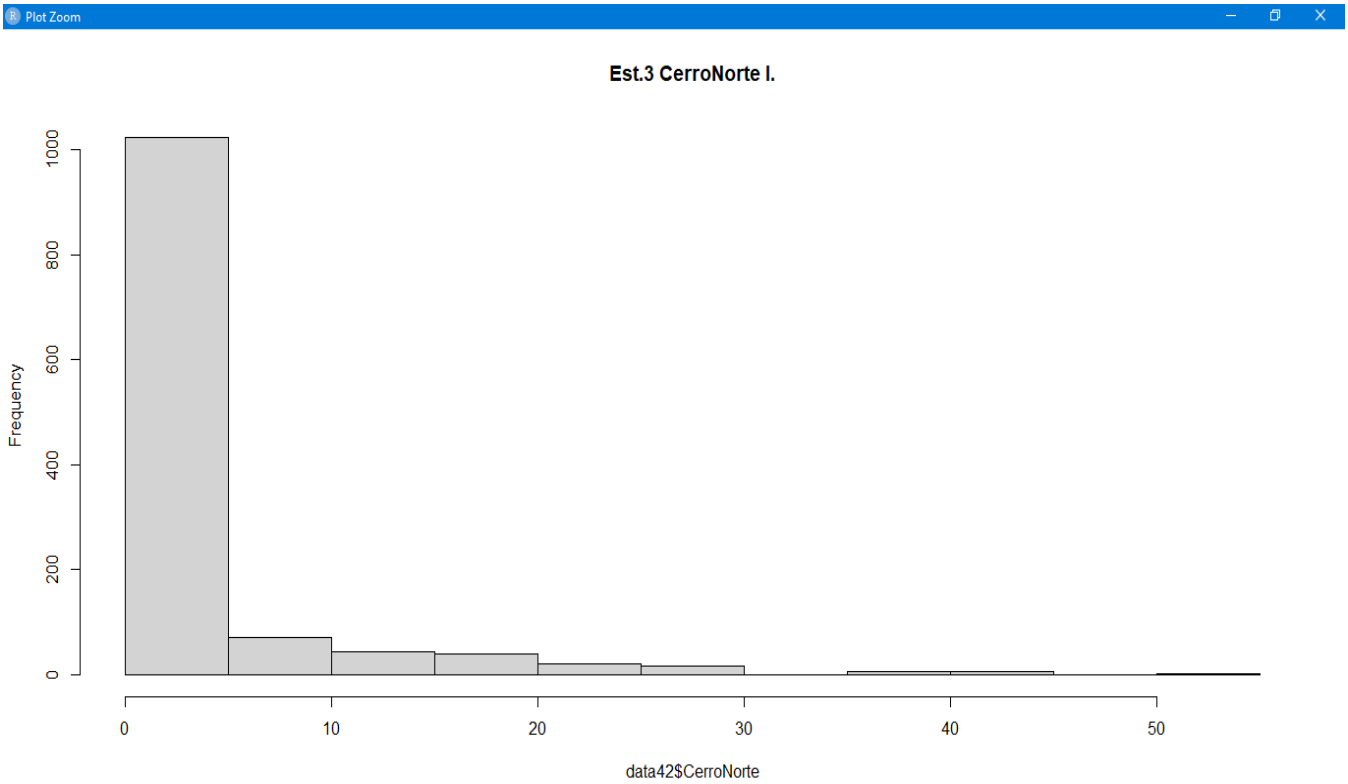
N.2 Estación Artillería



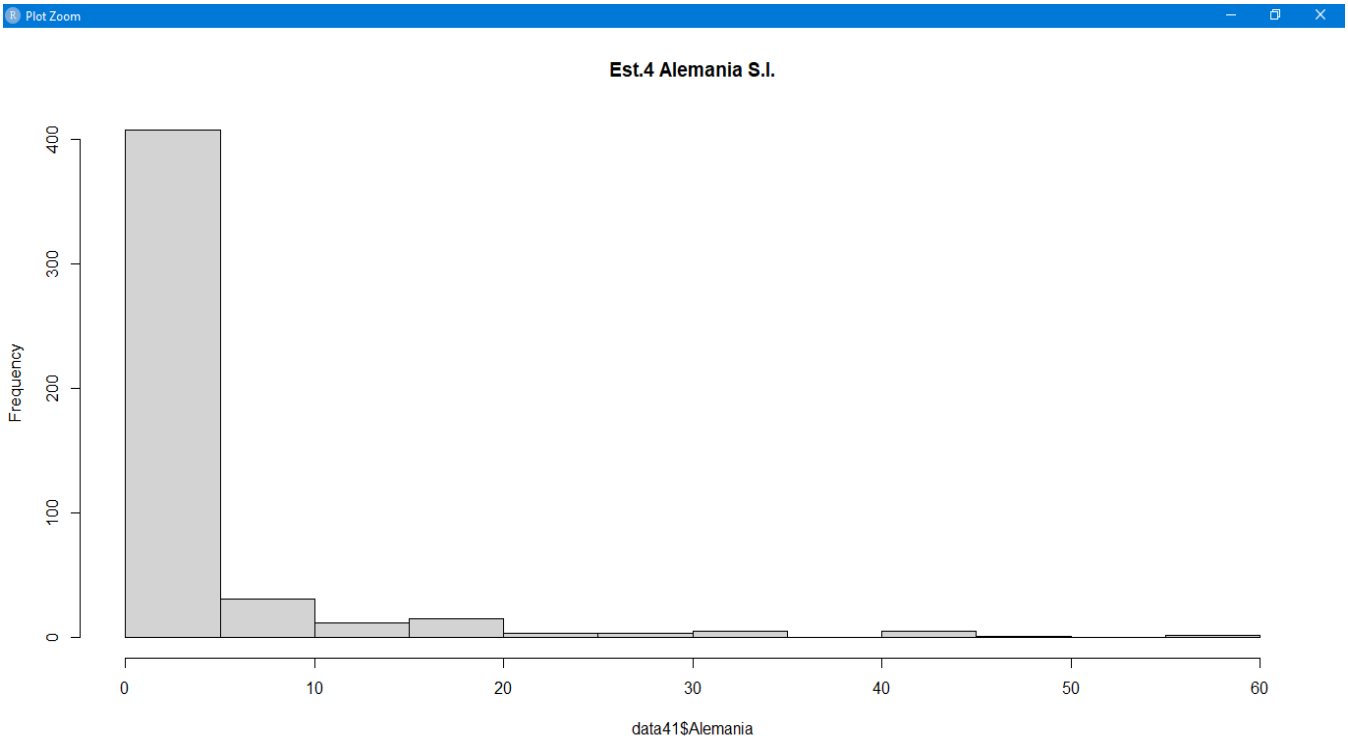


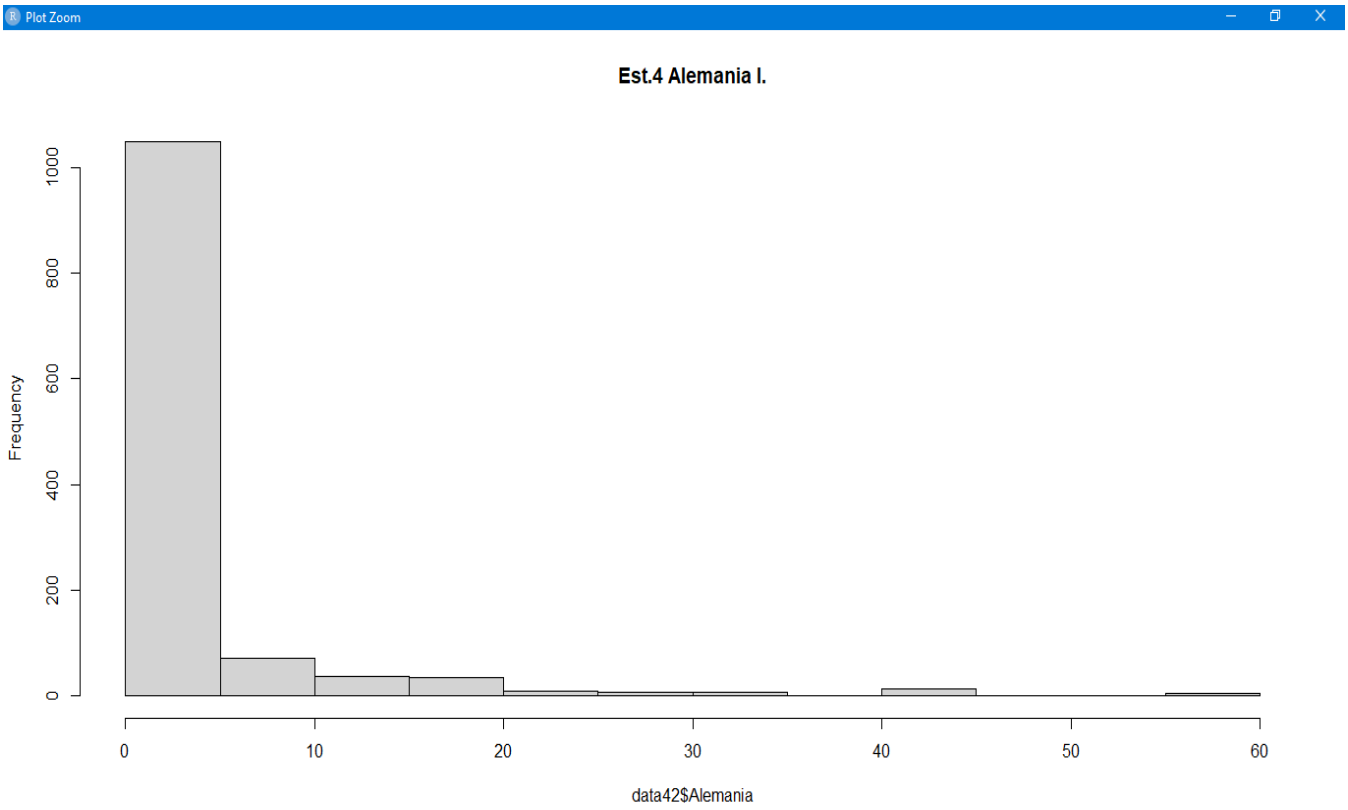
N.3 Estación CerroNorte



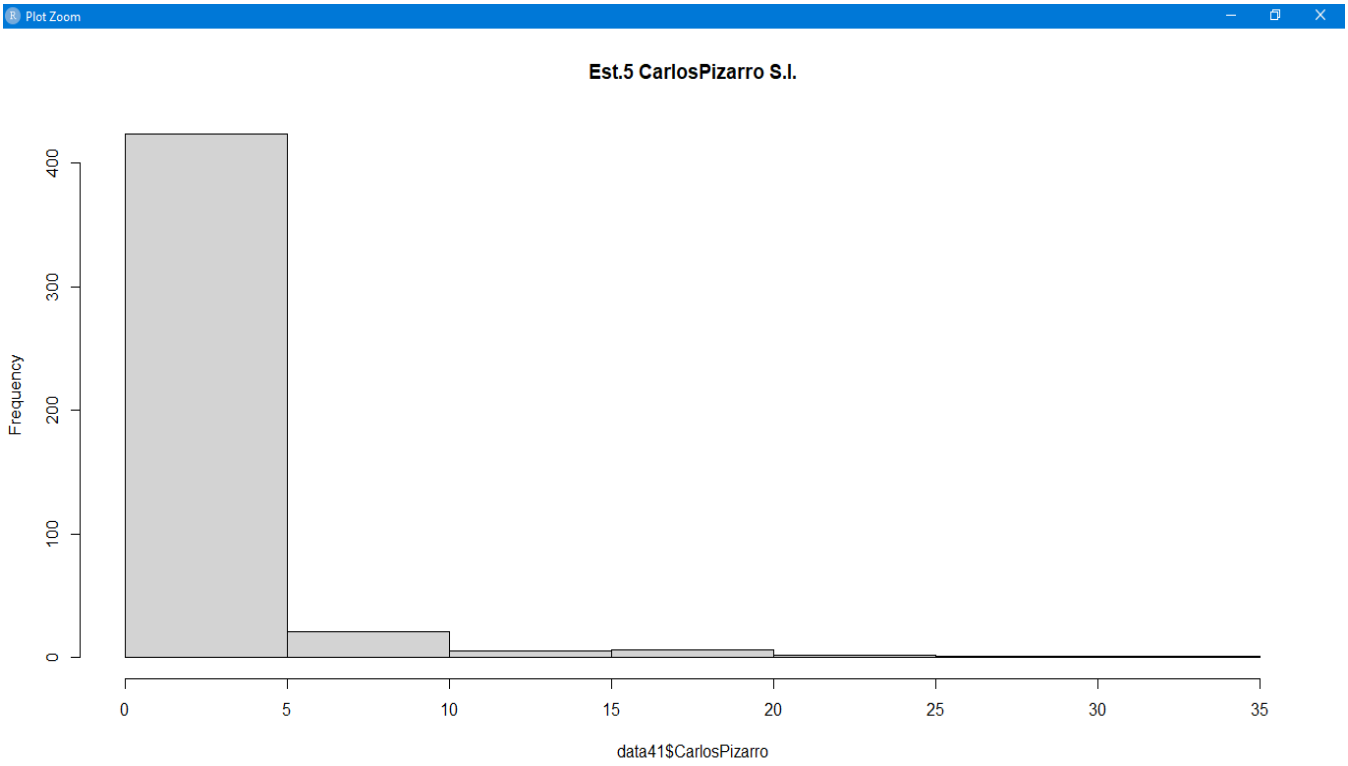


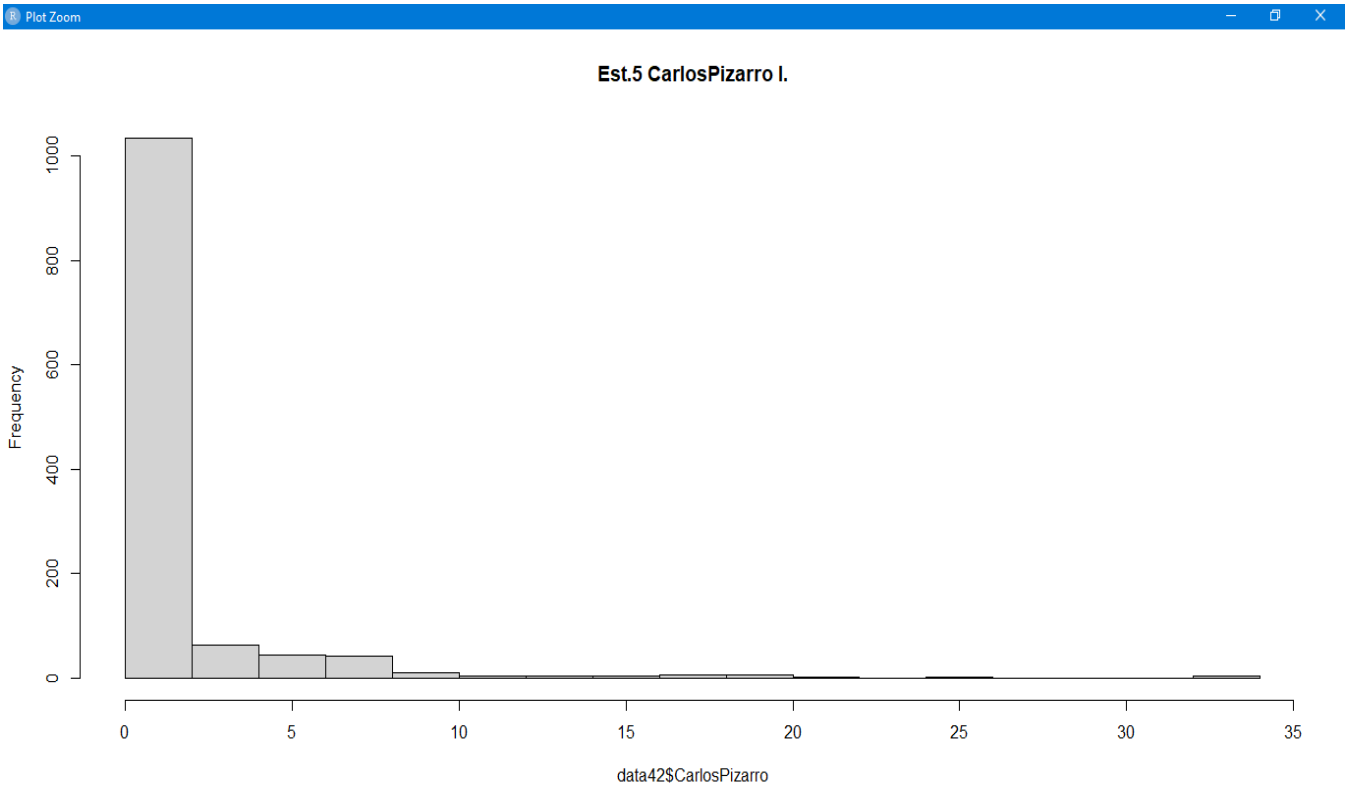
N.4 Estación Alemania



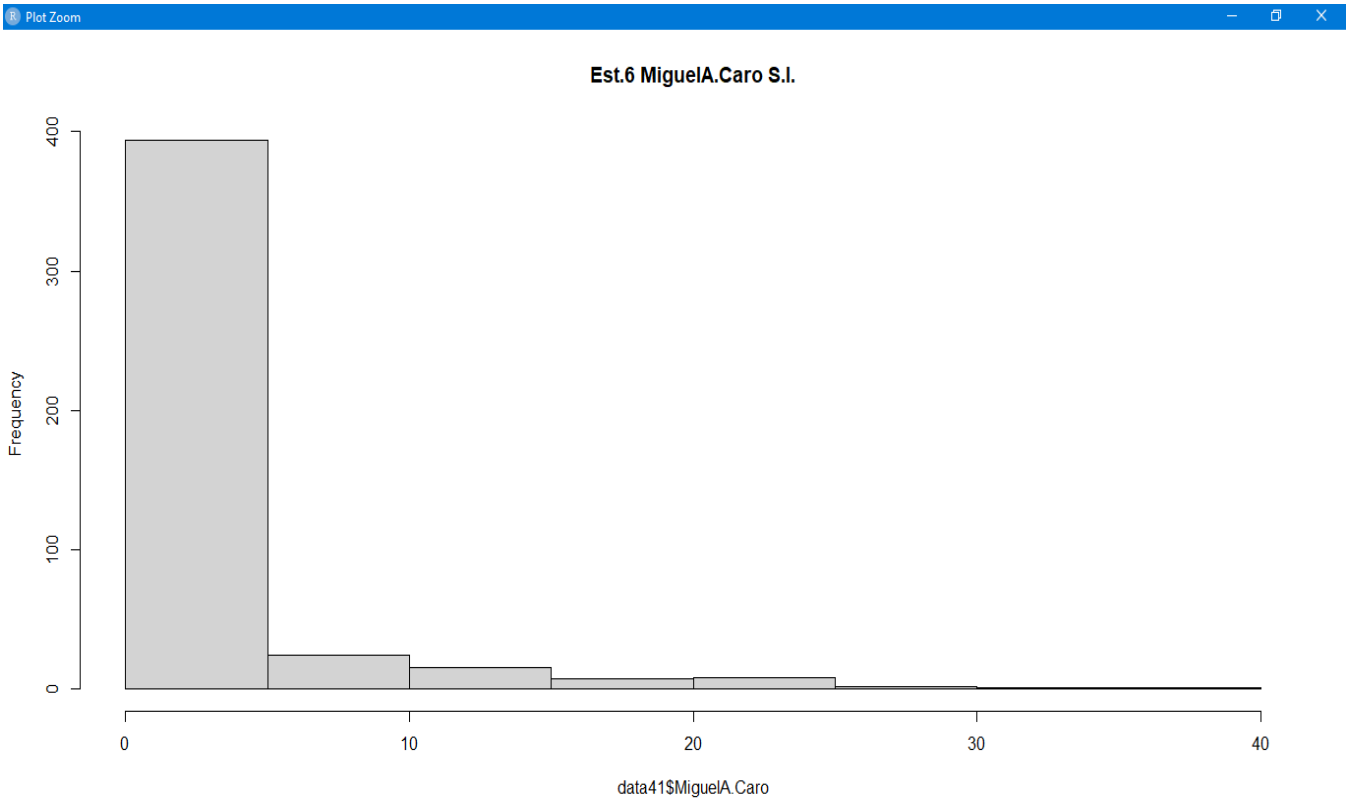


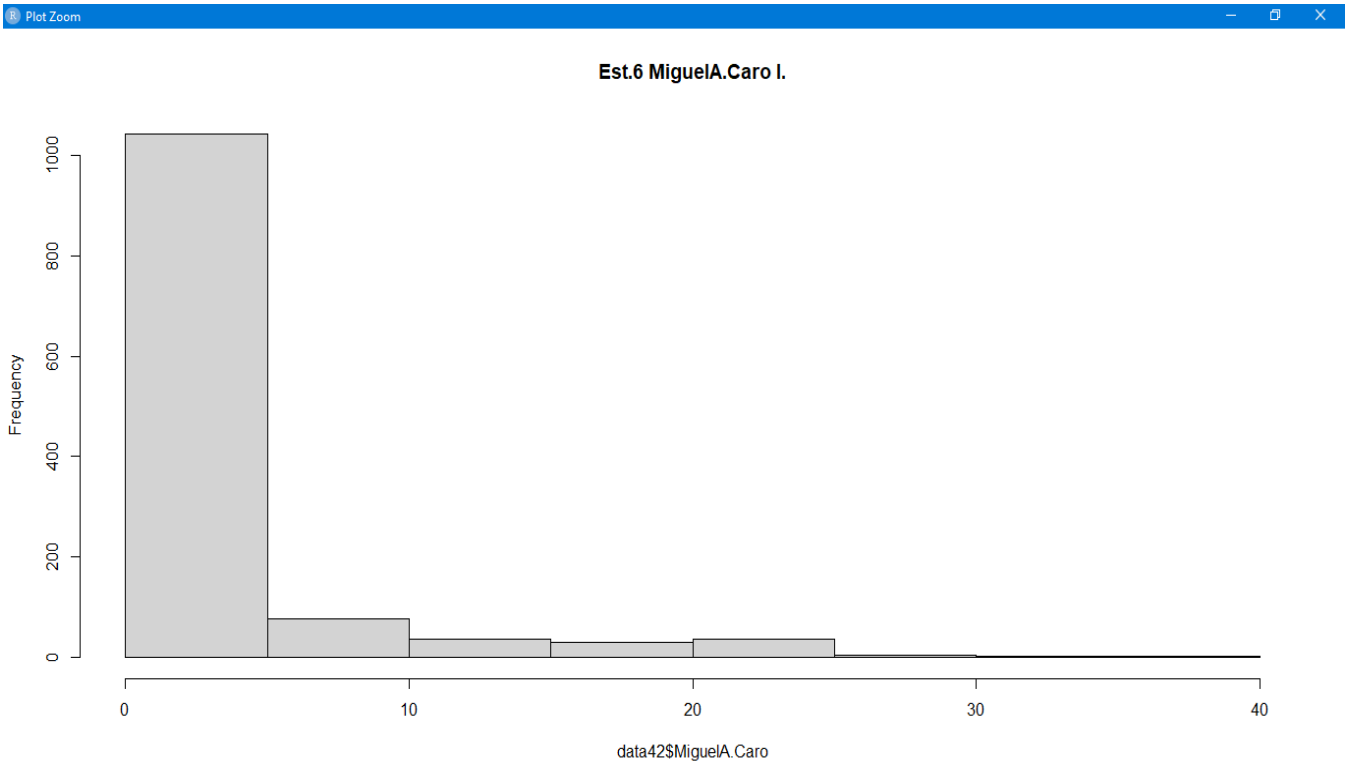
N.5 Estación CarlosPizarro



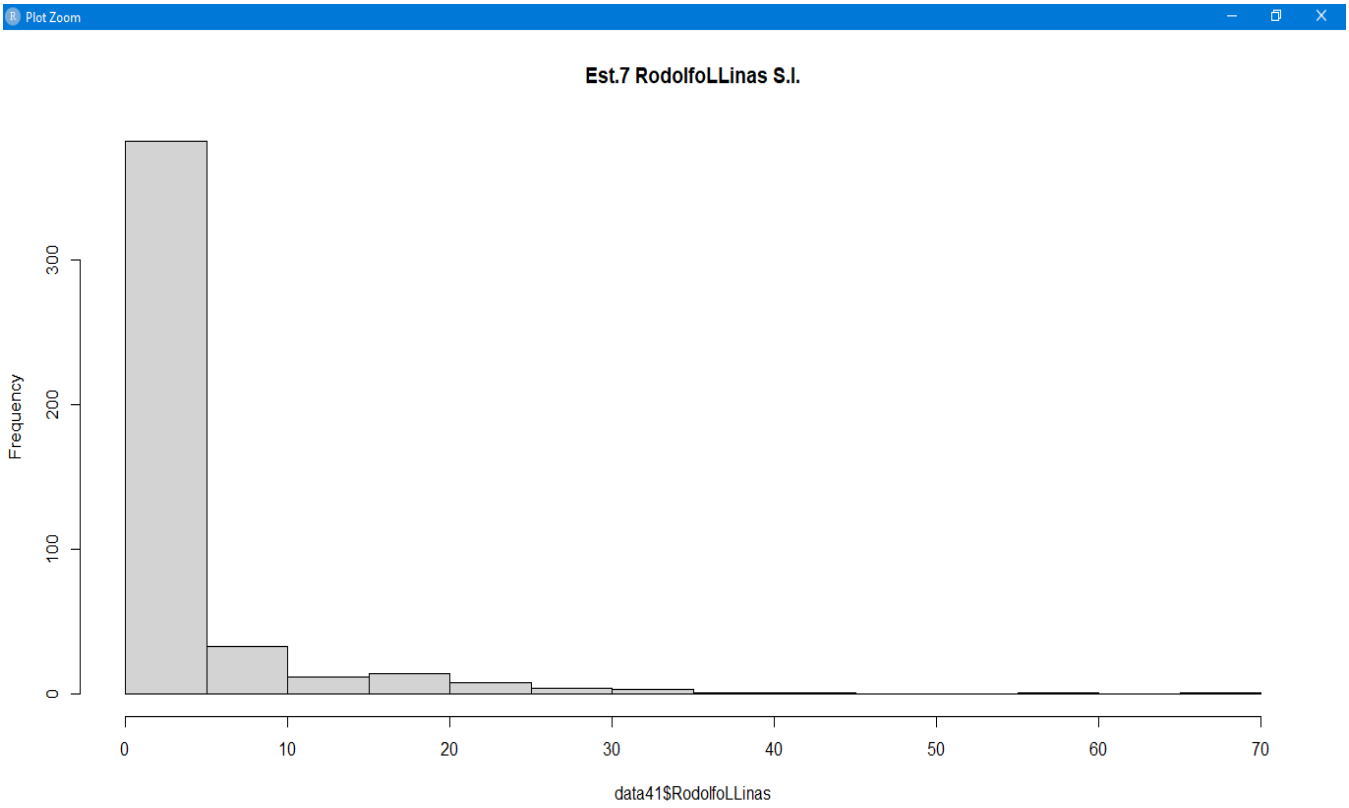


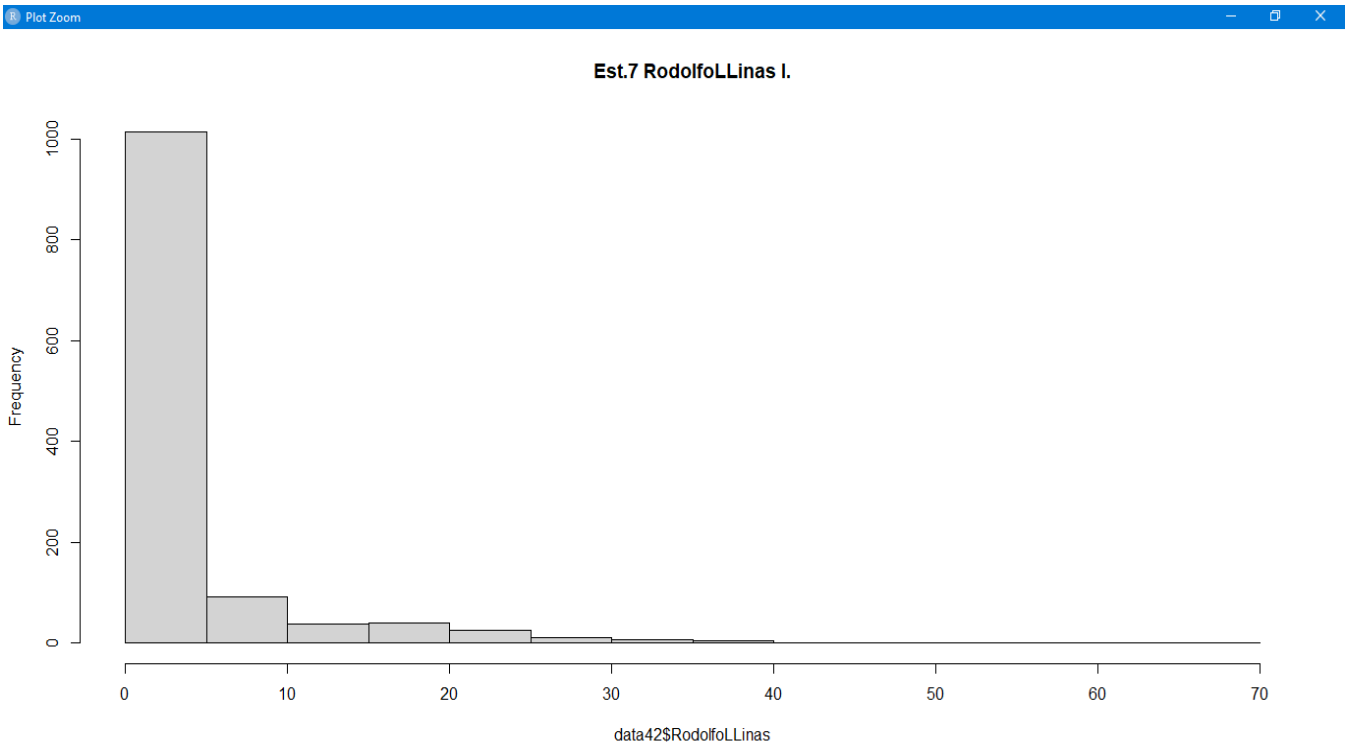
N.6 Estación MiguelACaro



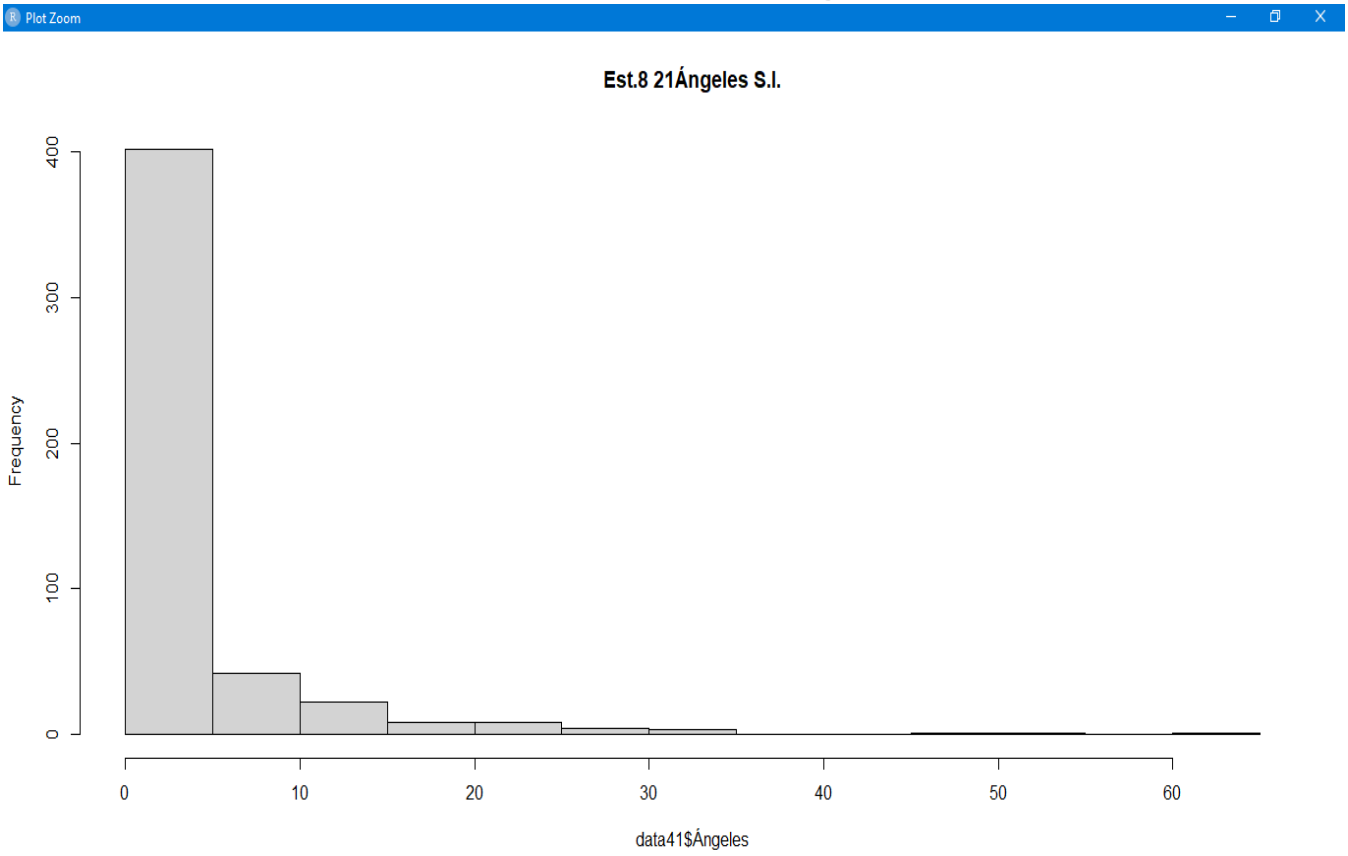


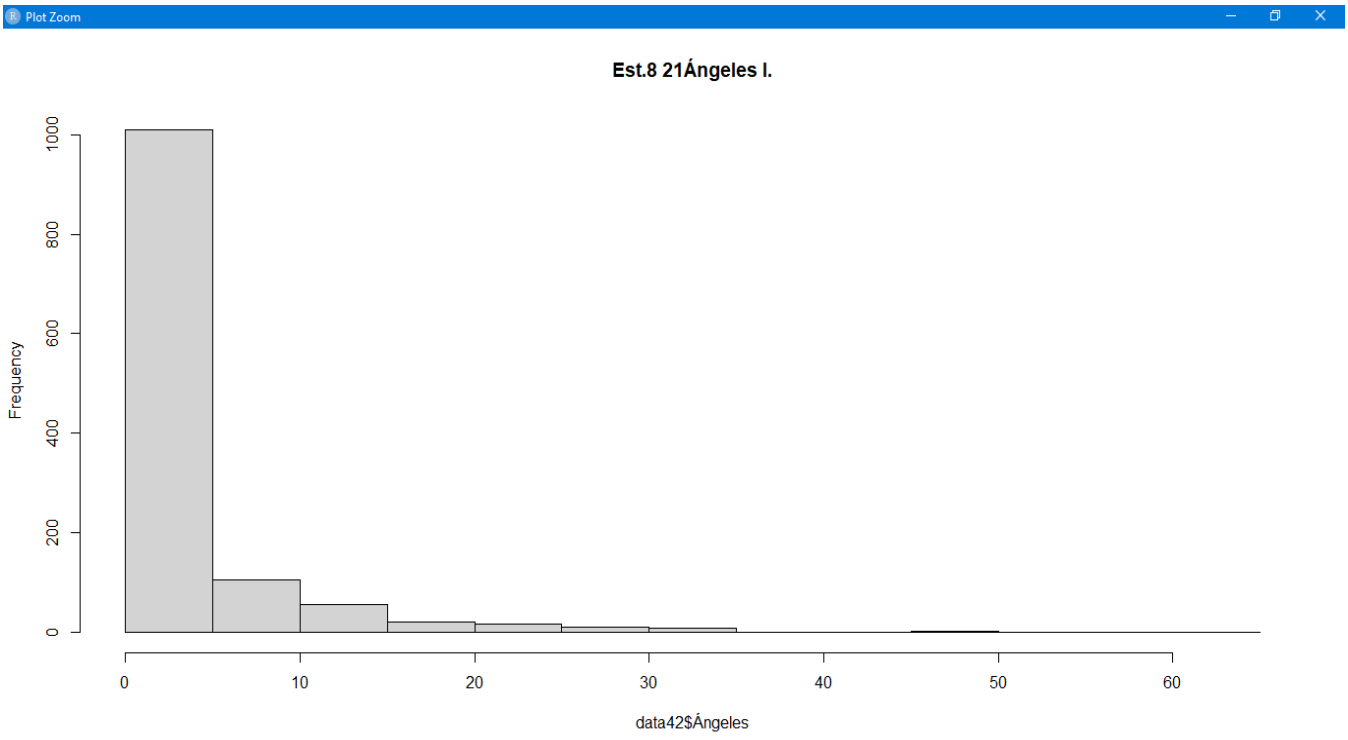
N. 7 Estación RodolfoLLinas



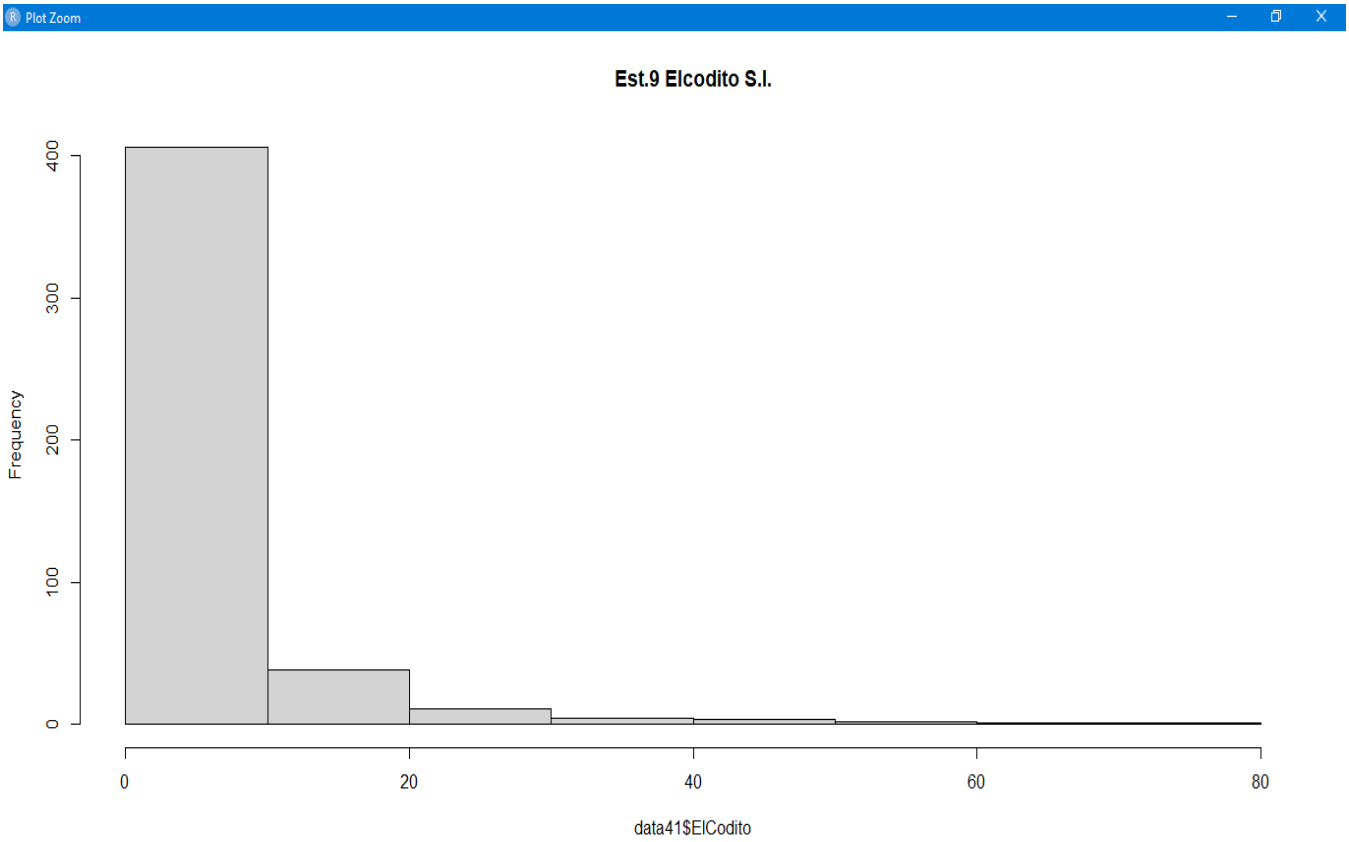


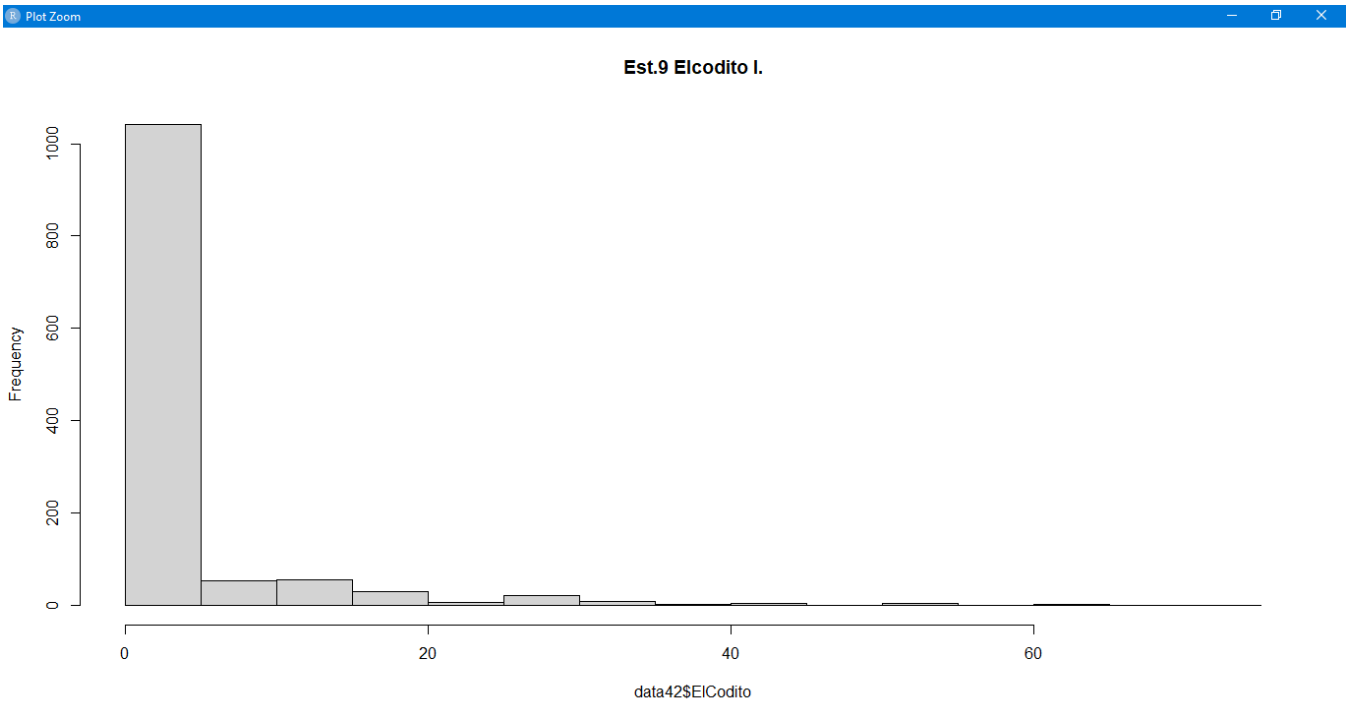
N. 8 Estación 21Ángeles



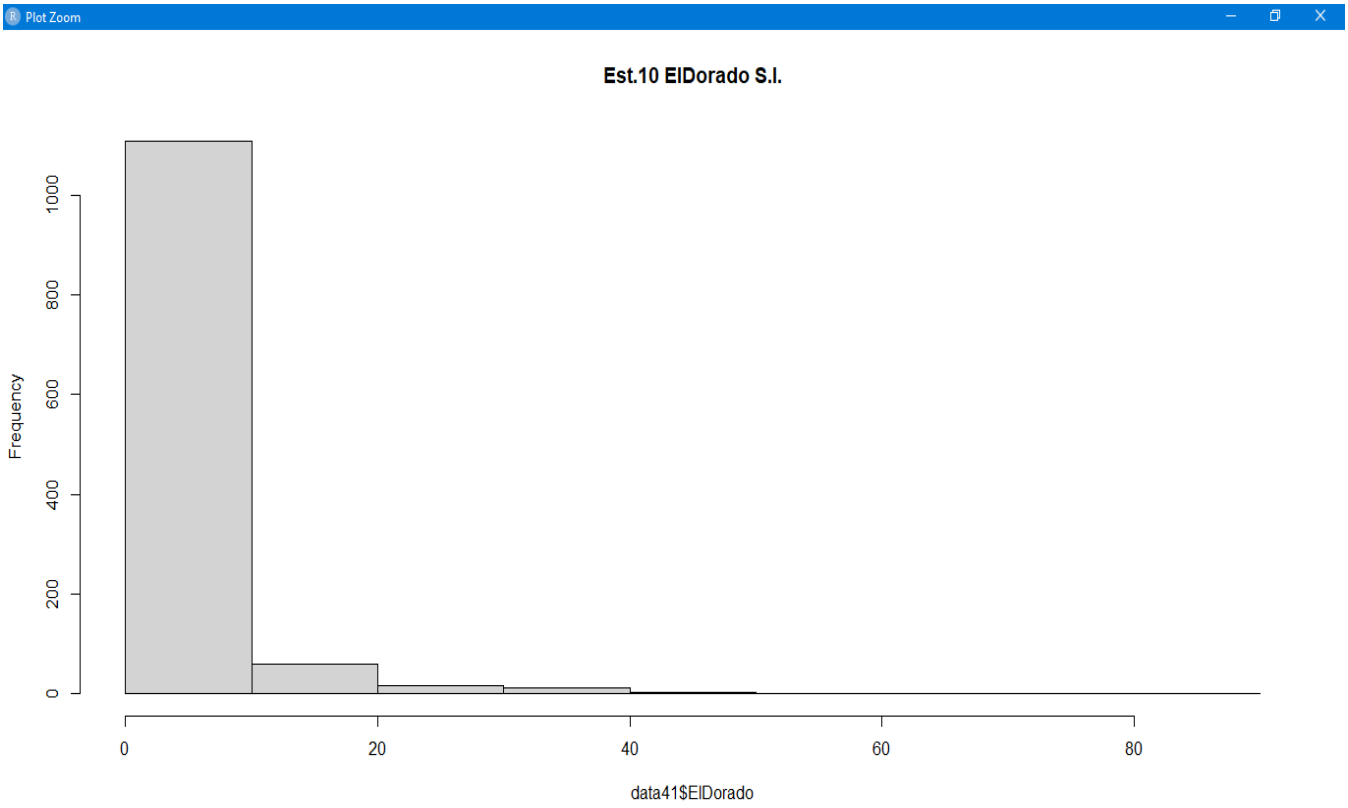


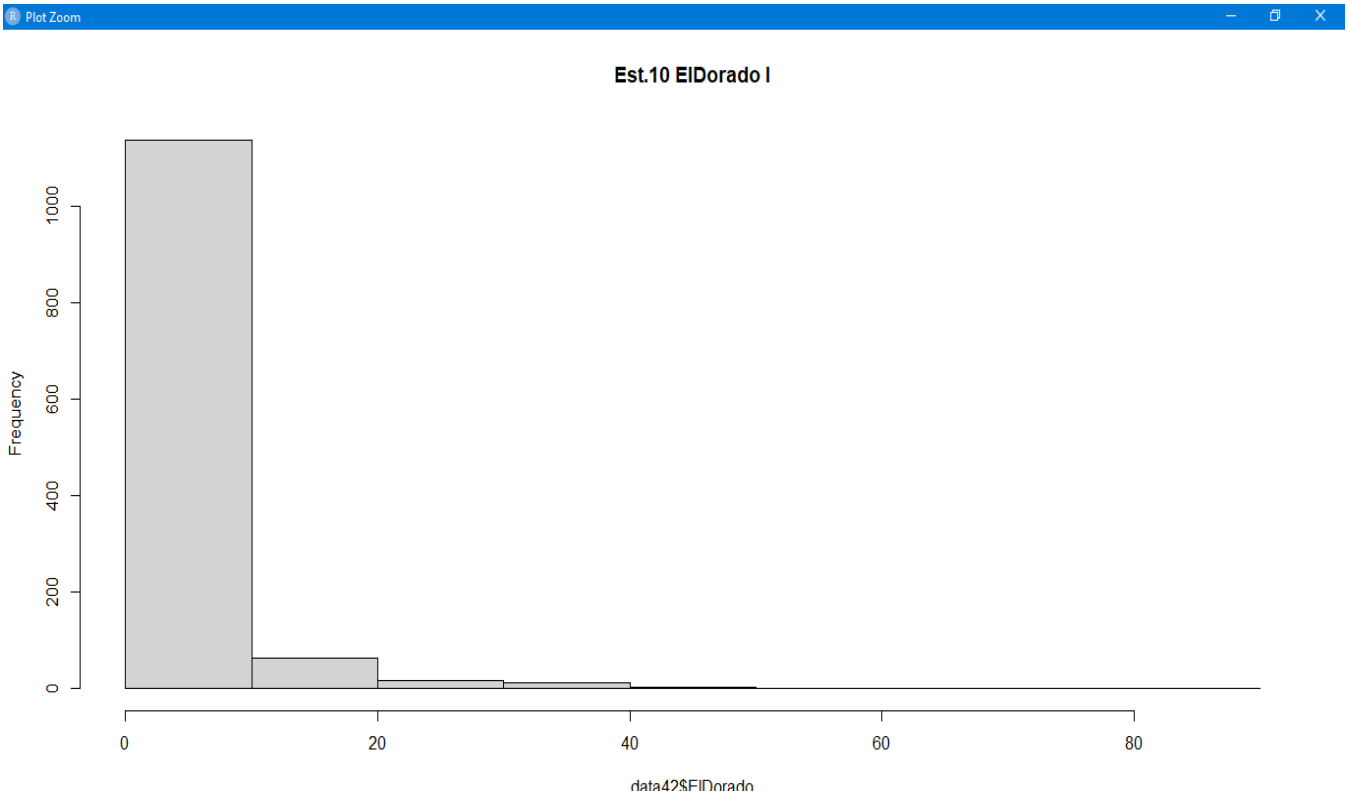
N. 9 Estación ElCodito



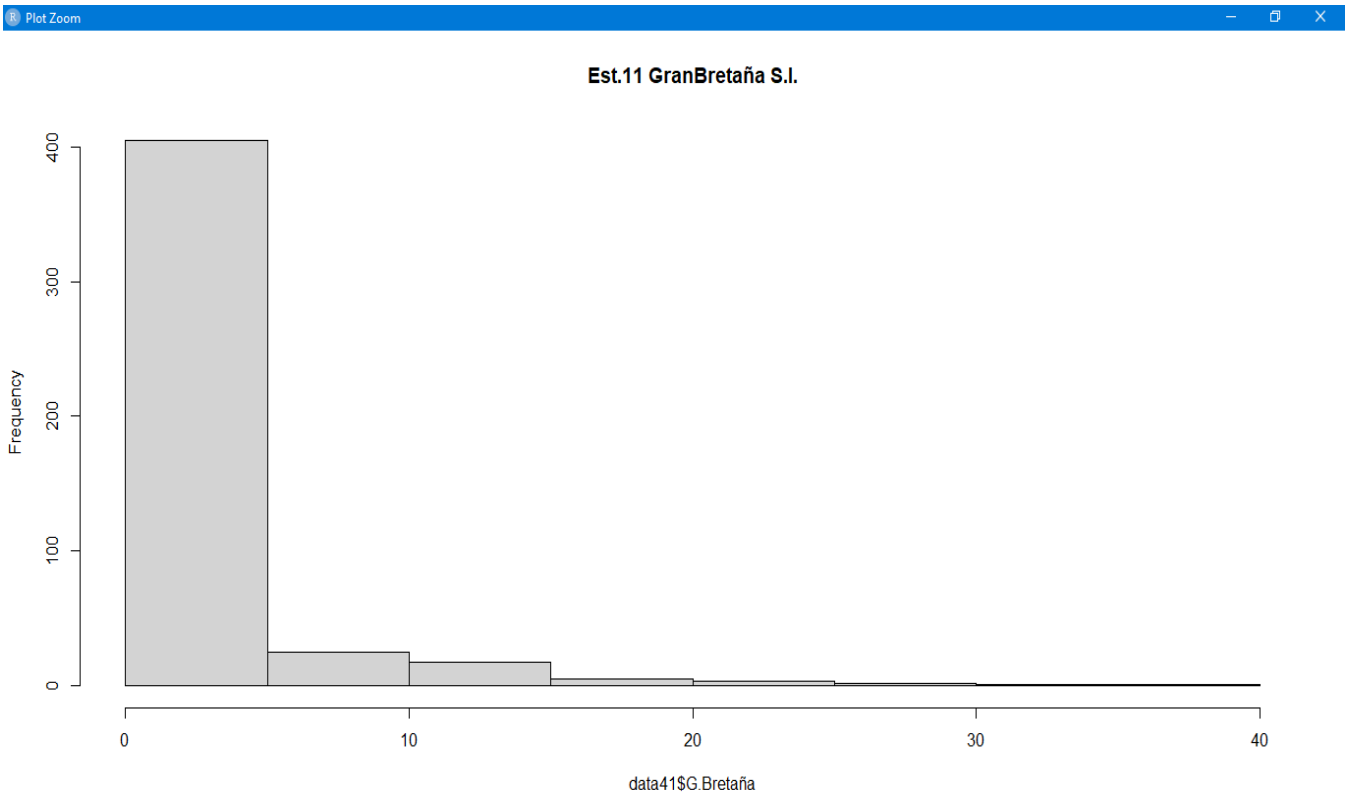


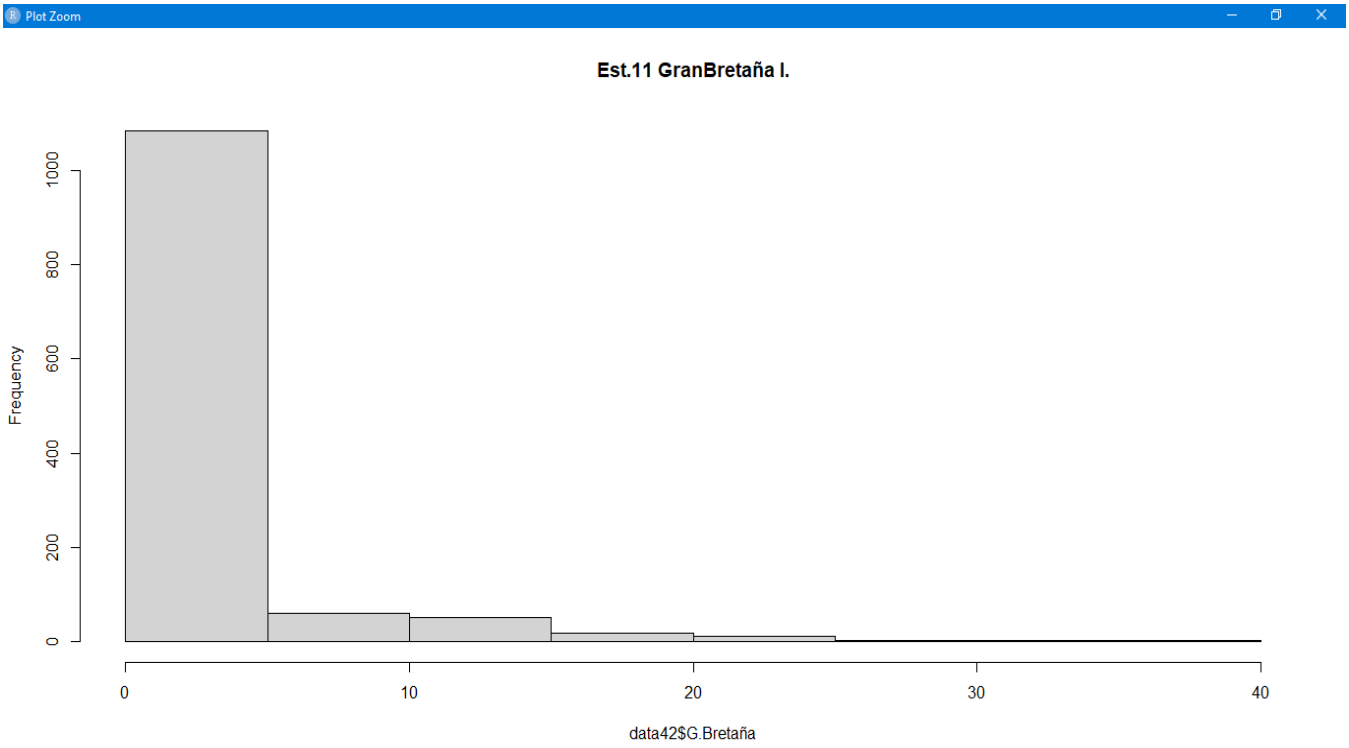
N. 10 Estación ElDorado



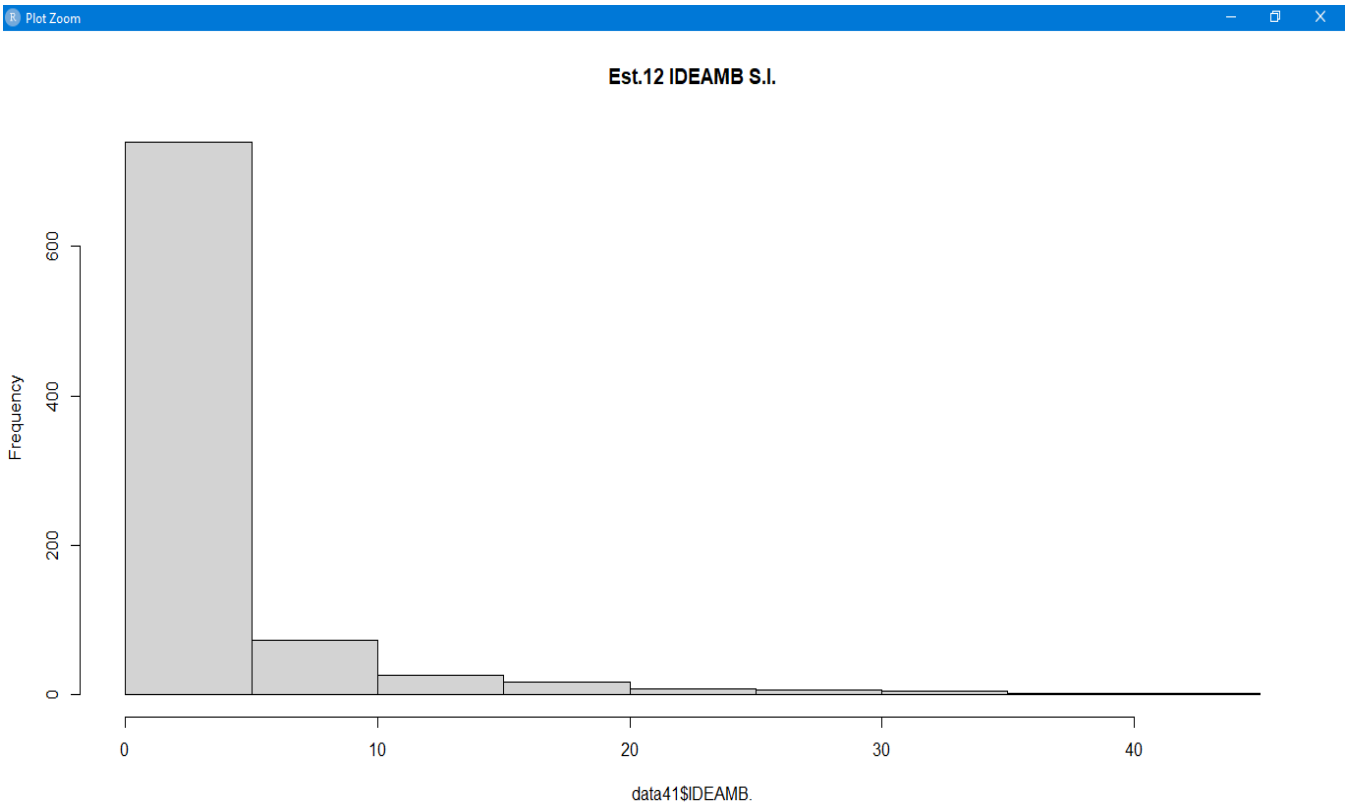


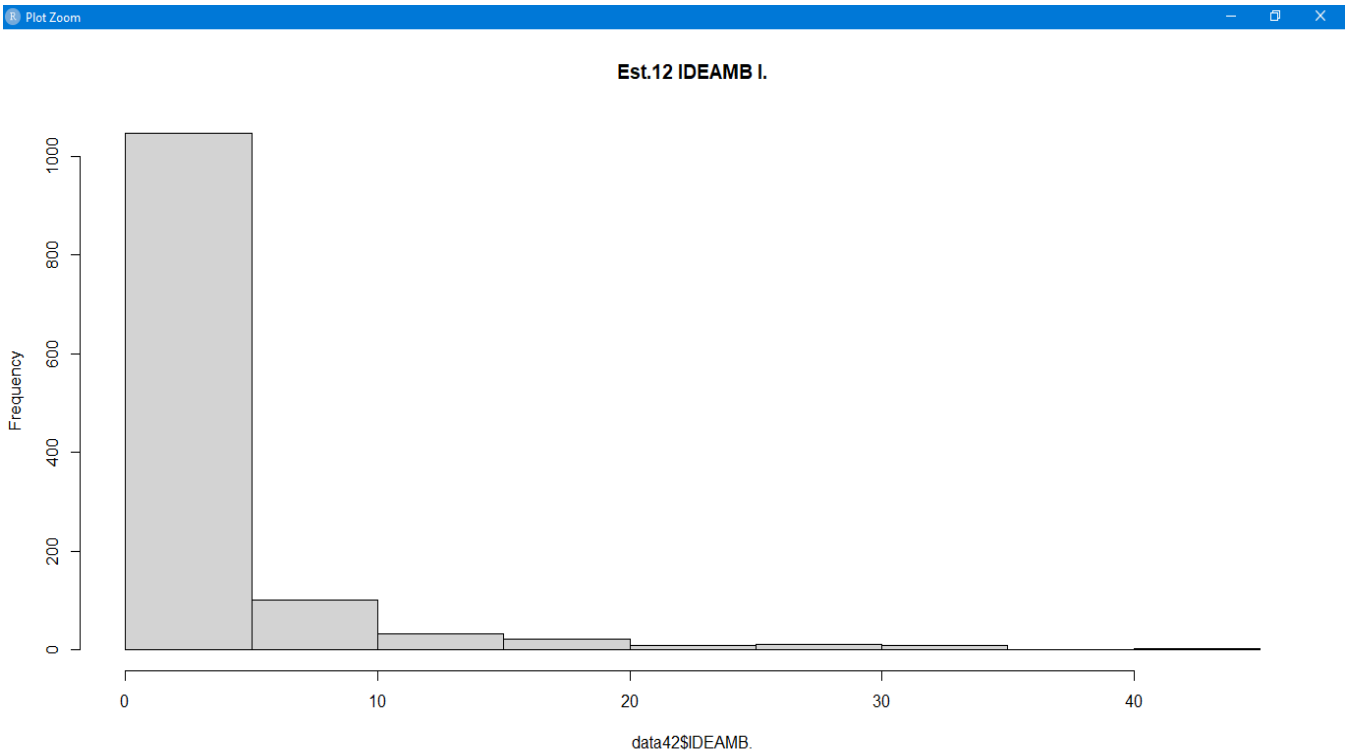
N. 11 Estación Bretaña



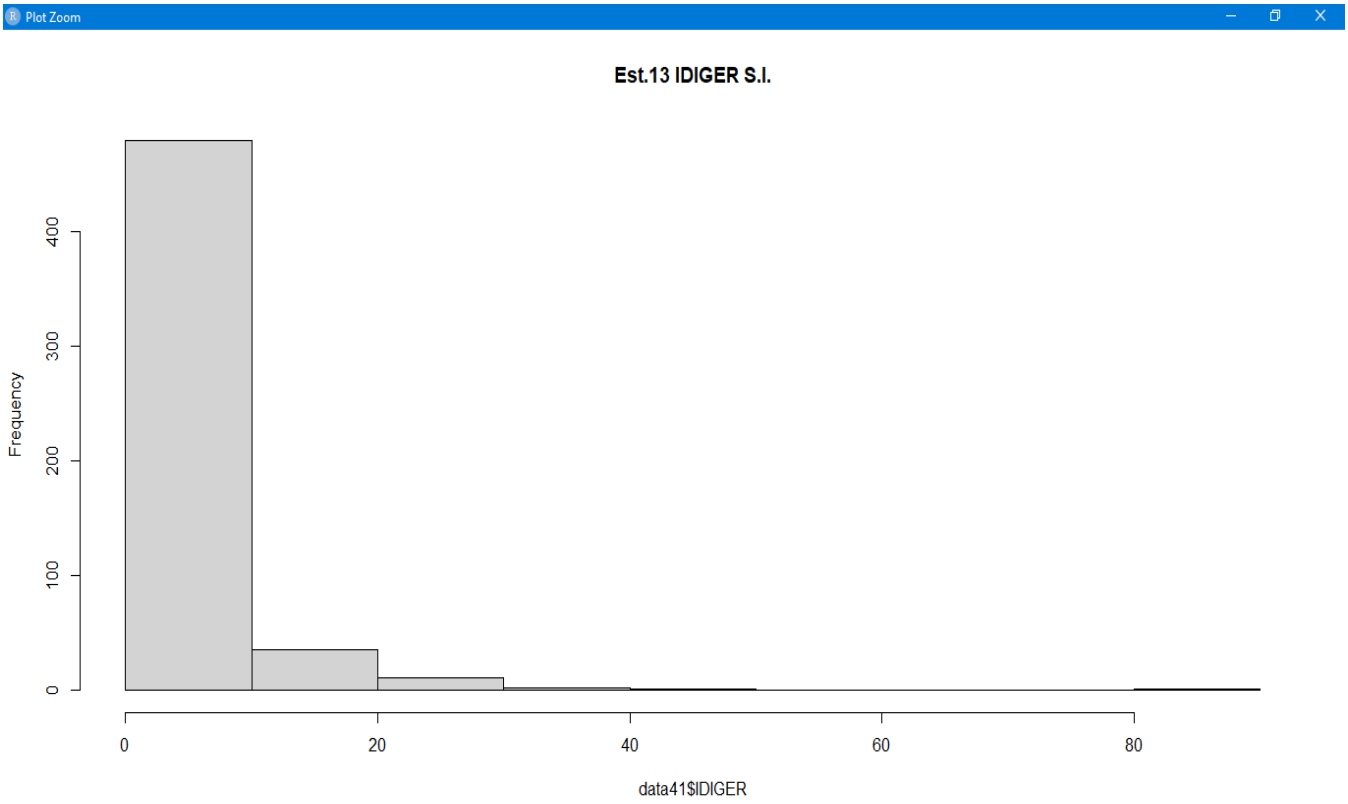


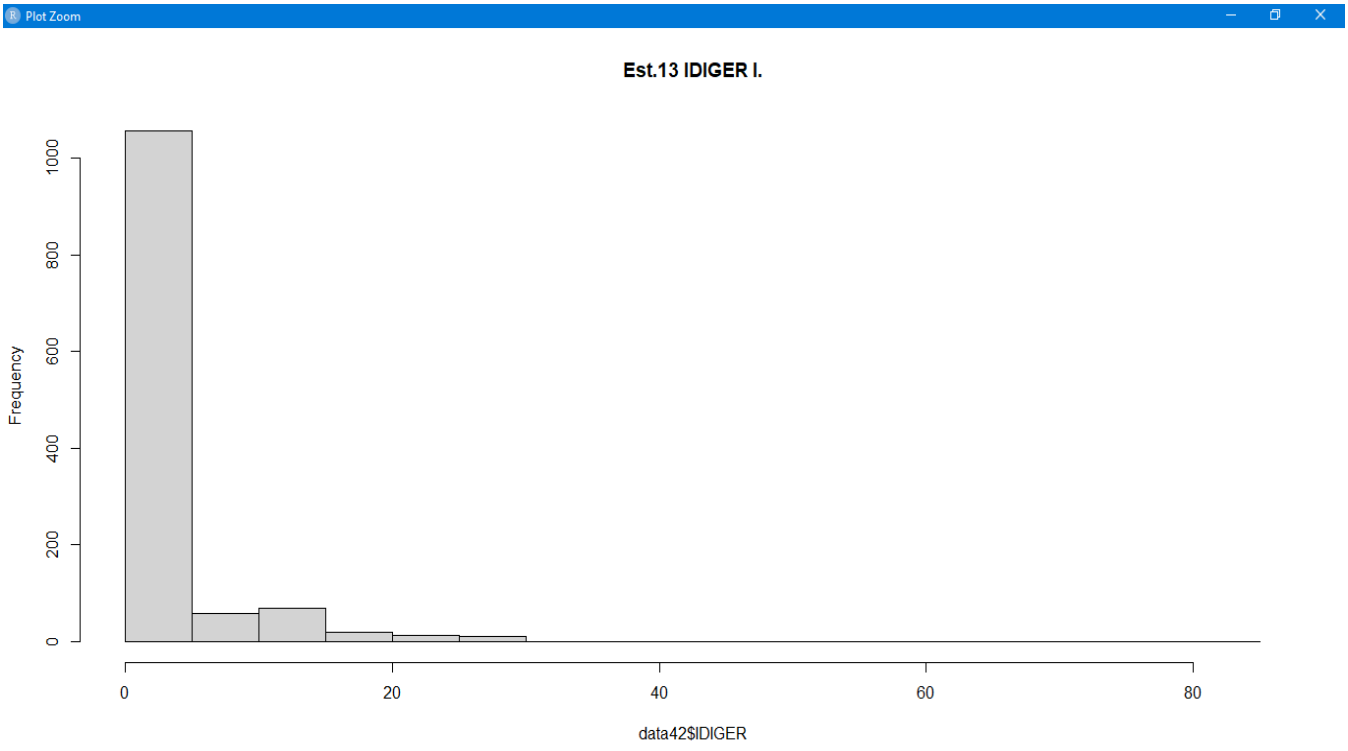
N. 12 Estación IDEAMB



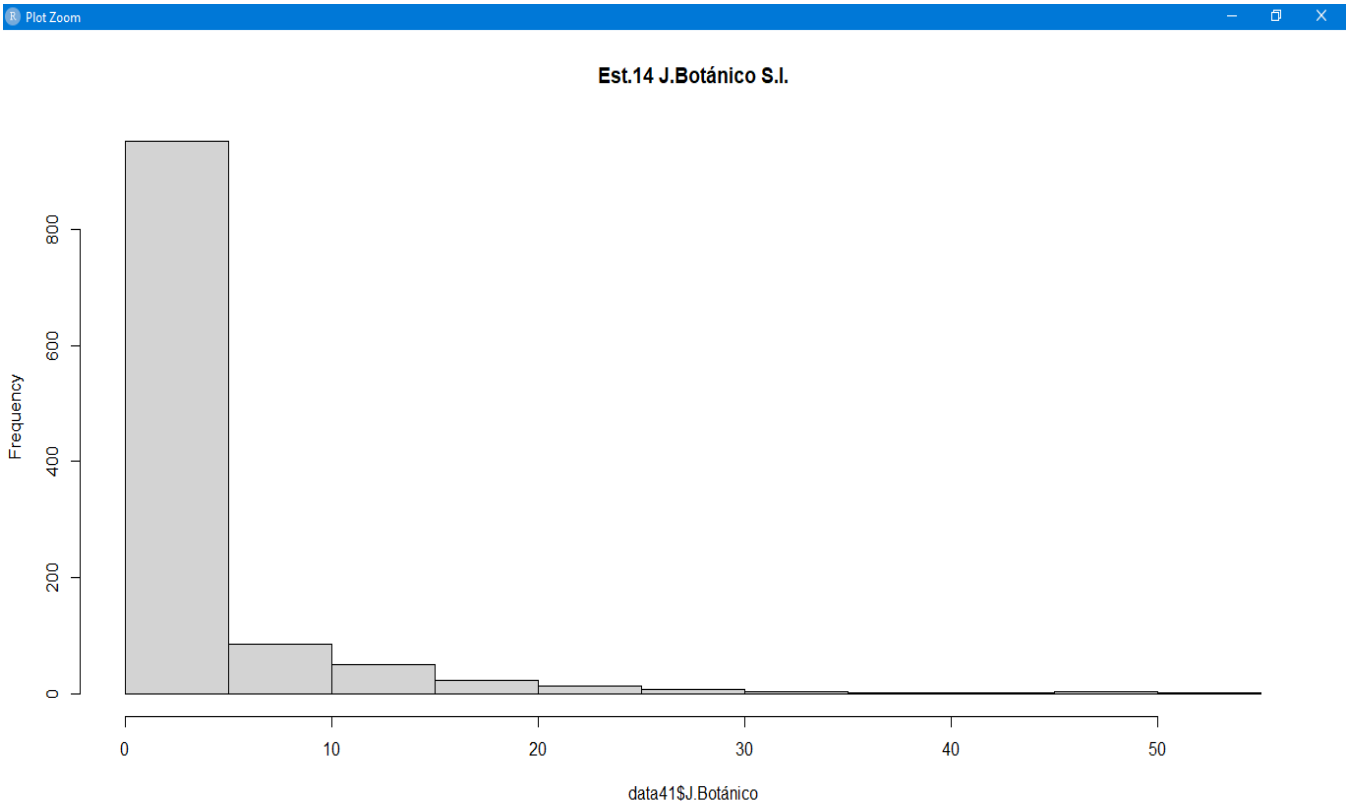


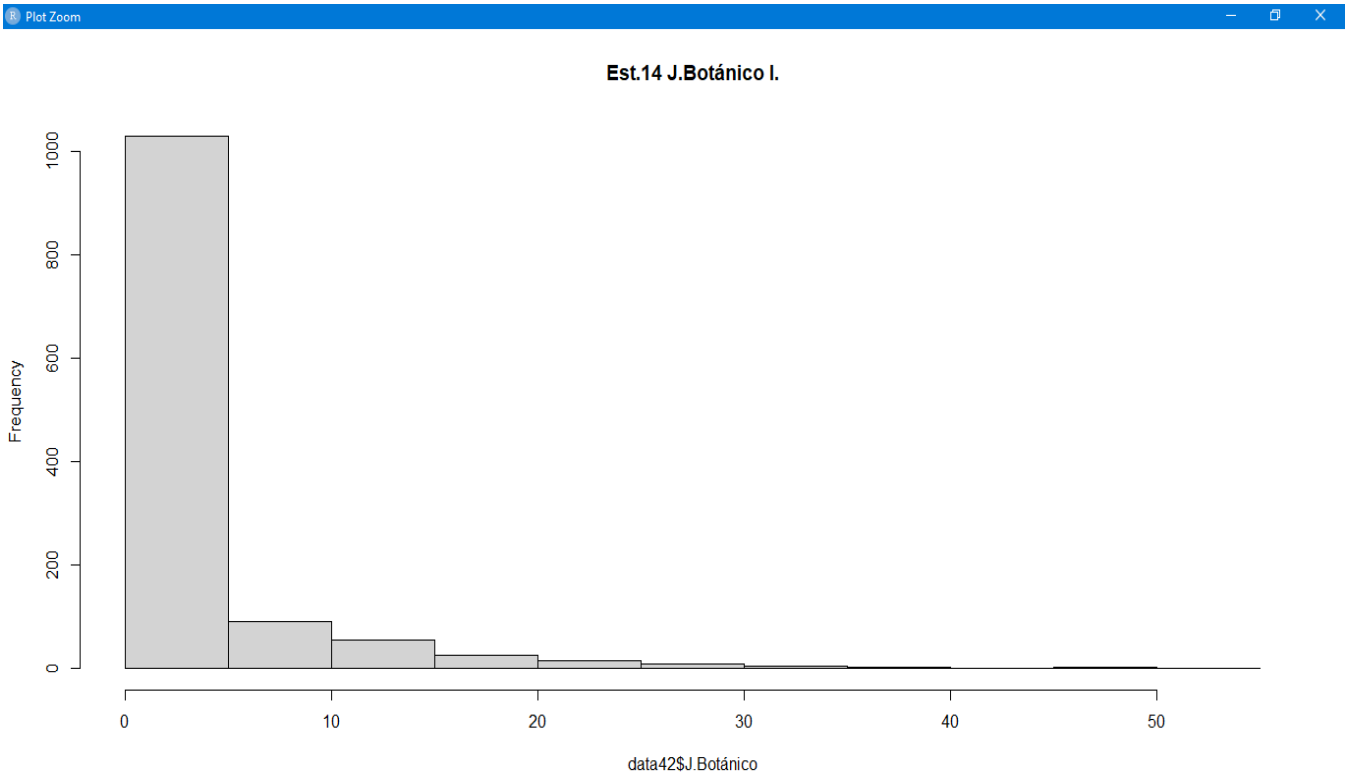
N. 13 Estación IDIGER



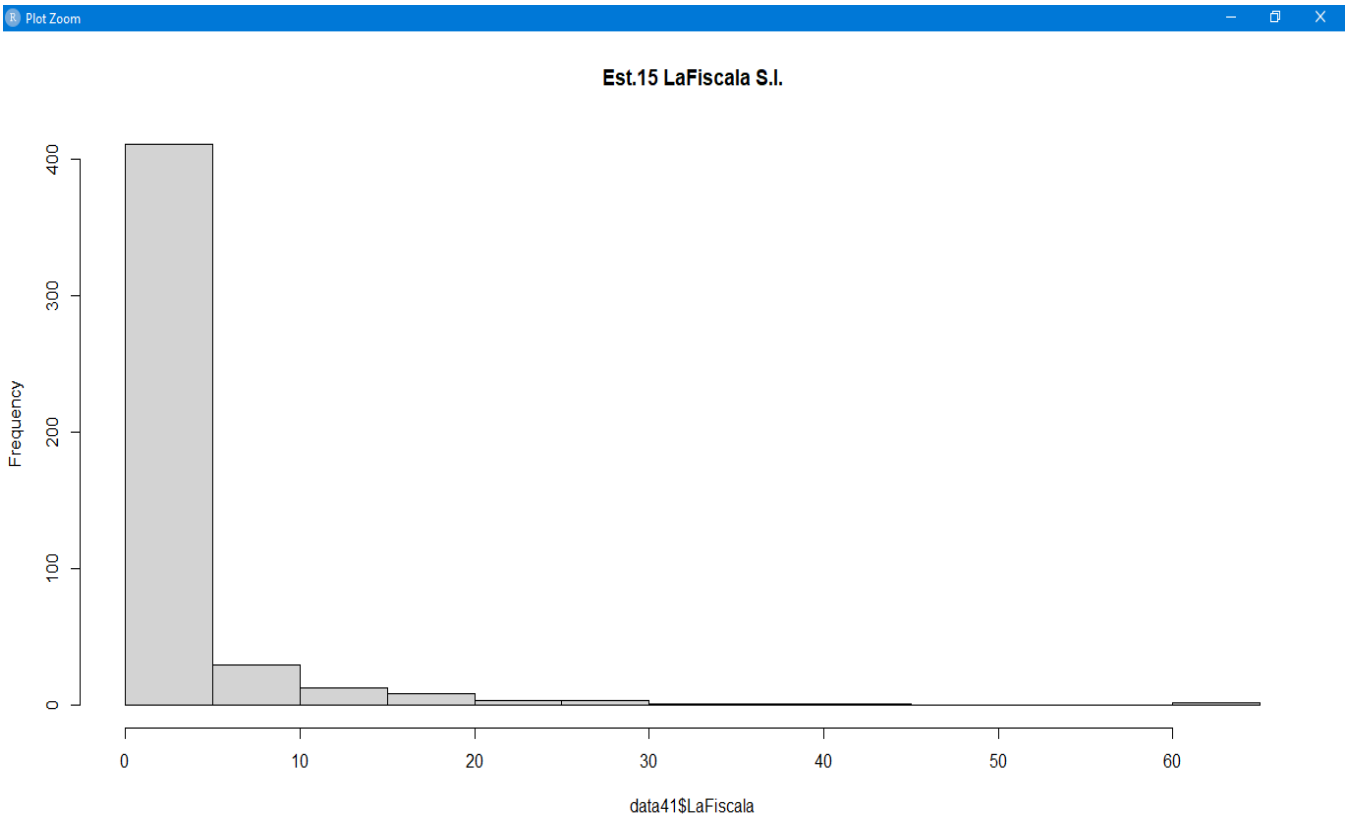


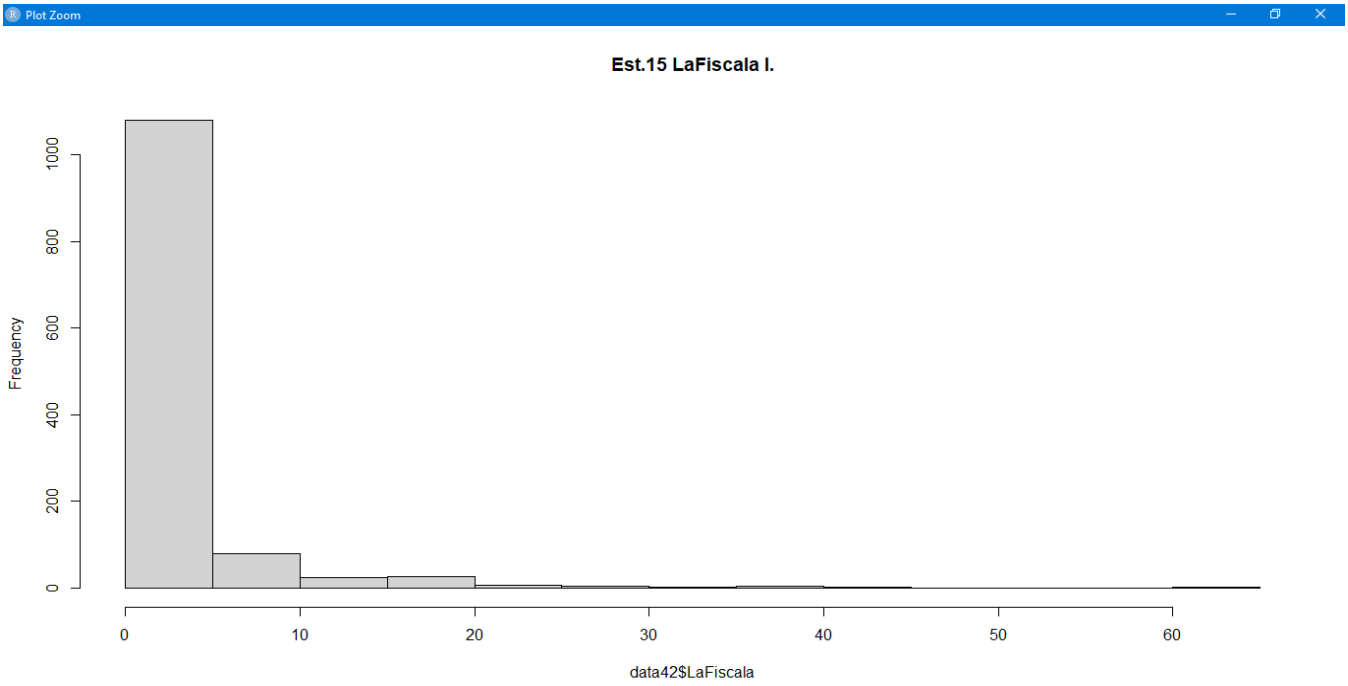
N. 14 Estación J.Botánico



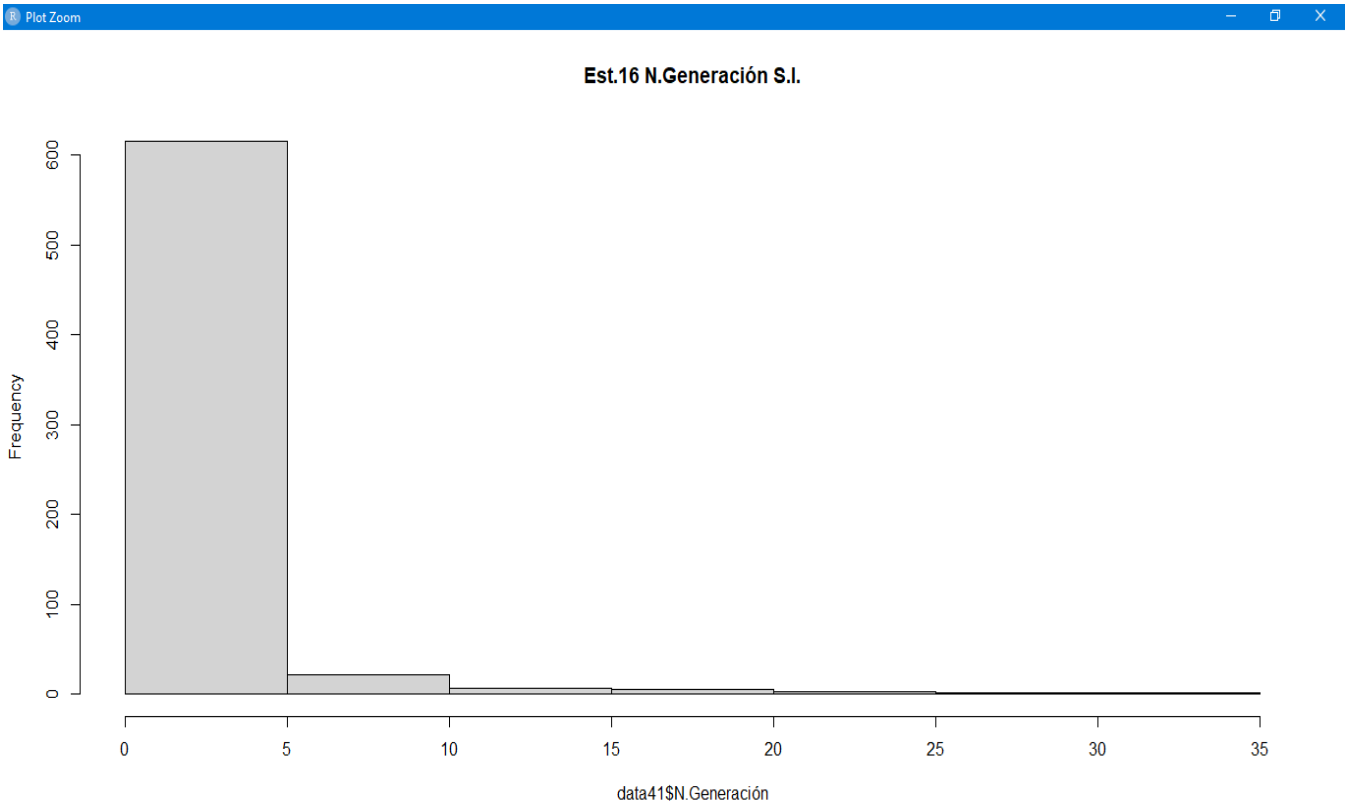


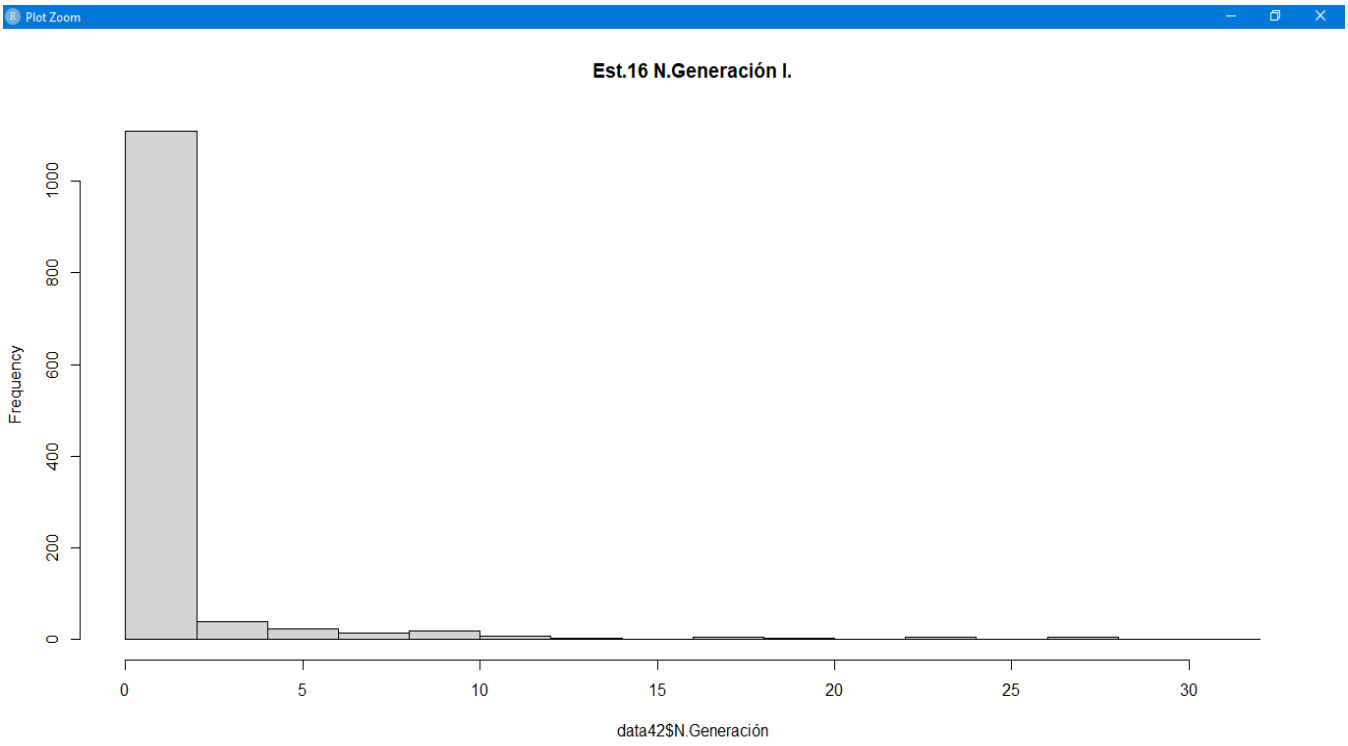
N. 15 Estación LaFiscala



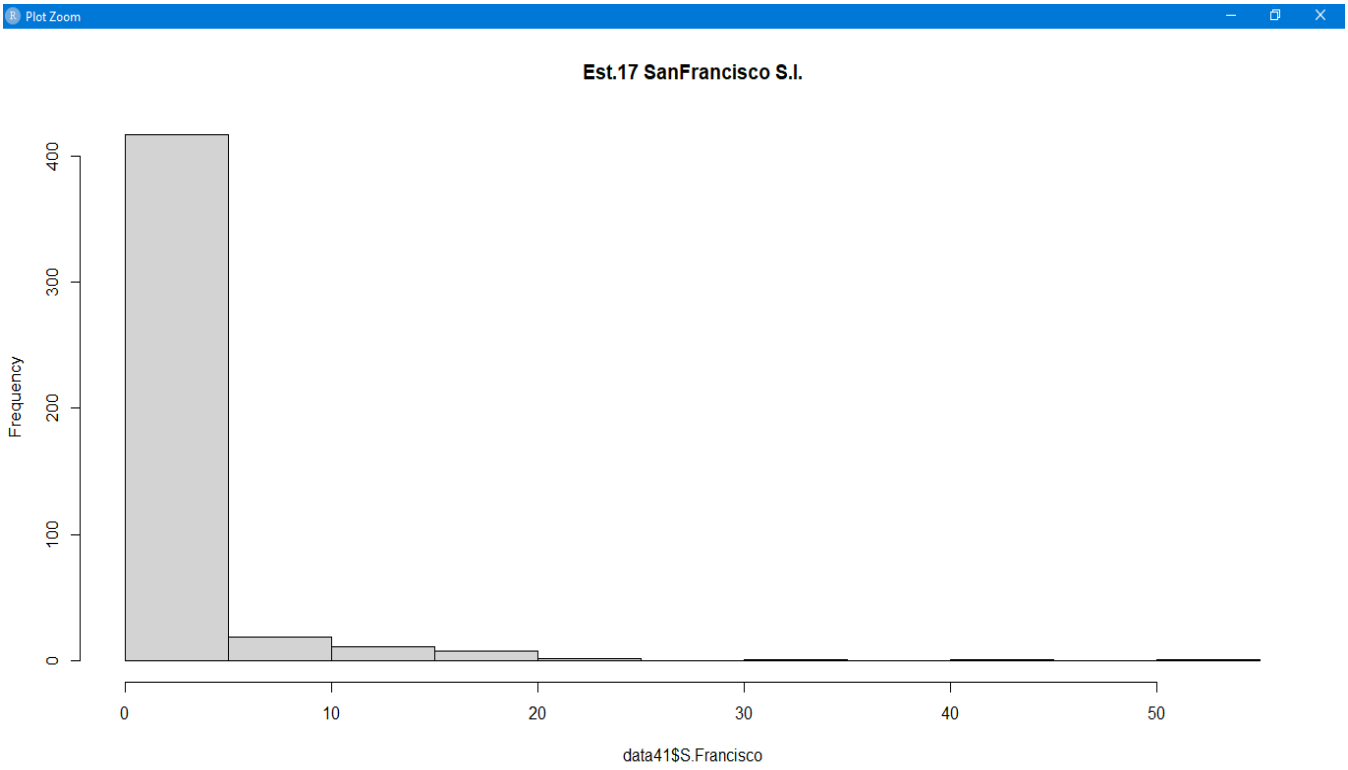


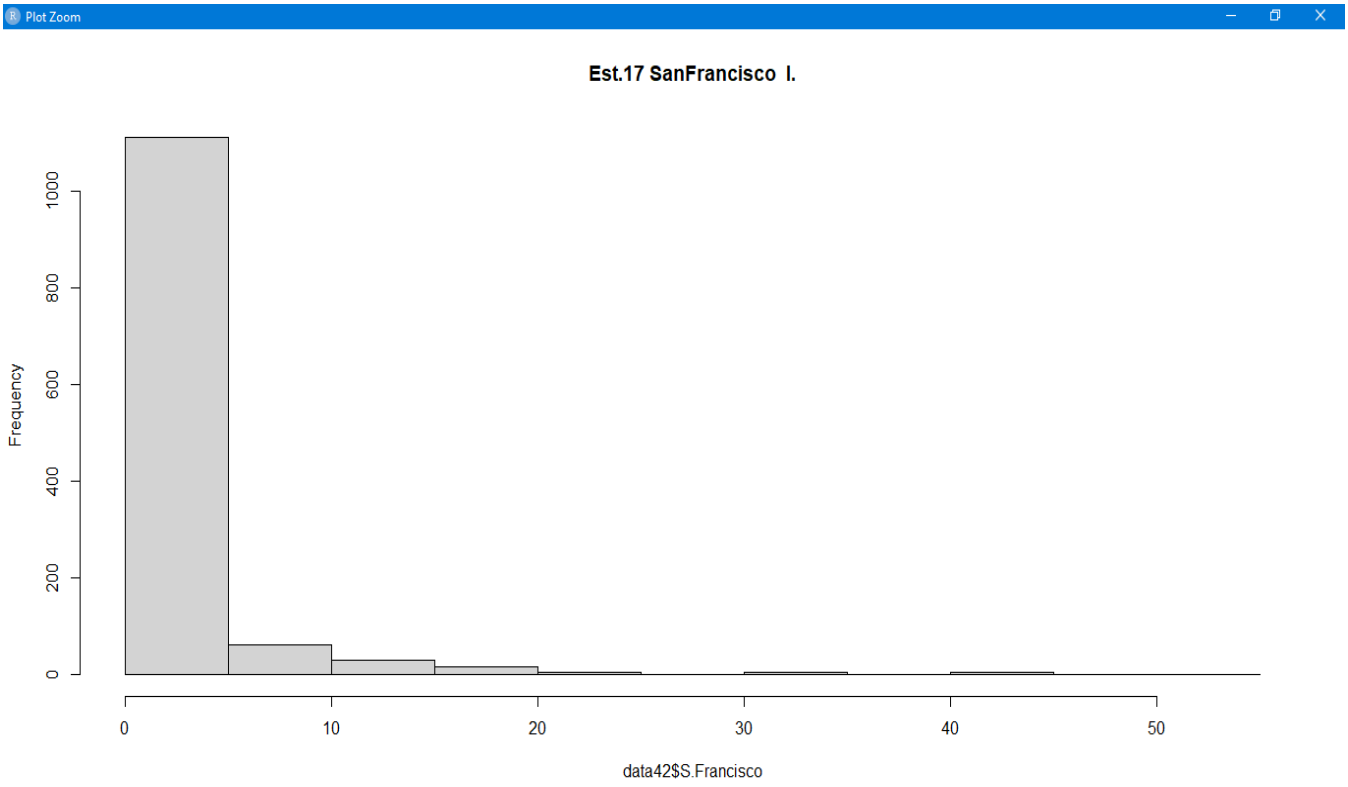
N. 16 Estación N. Generación



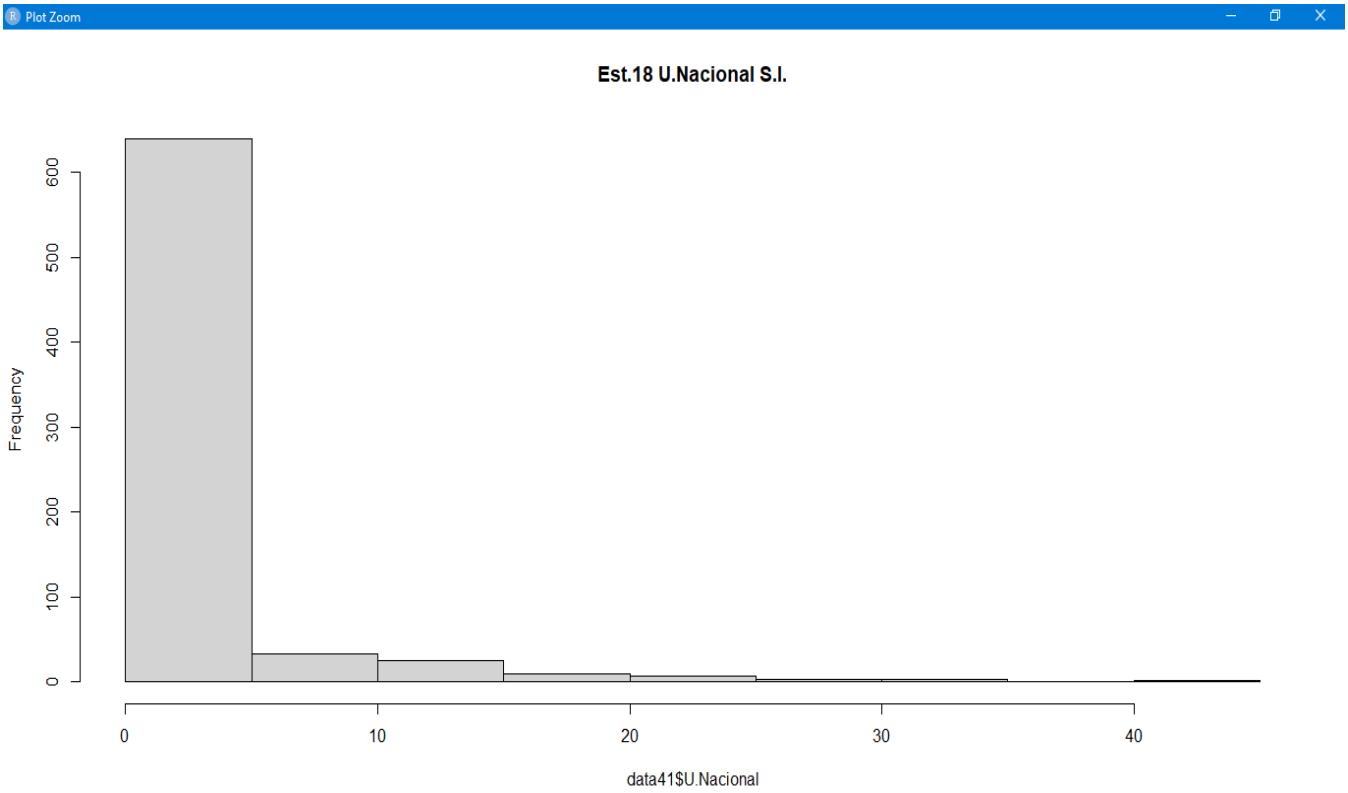


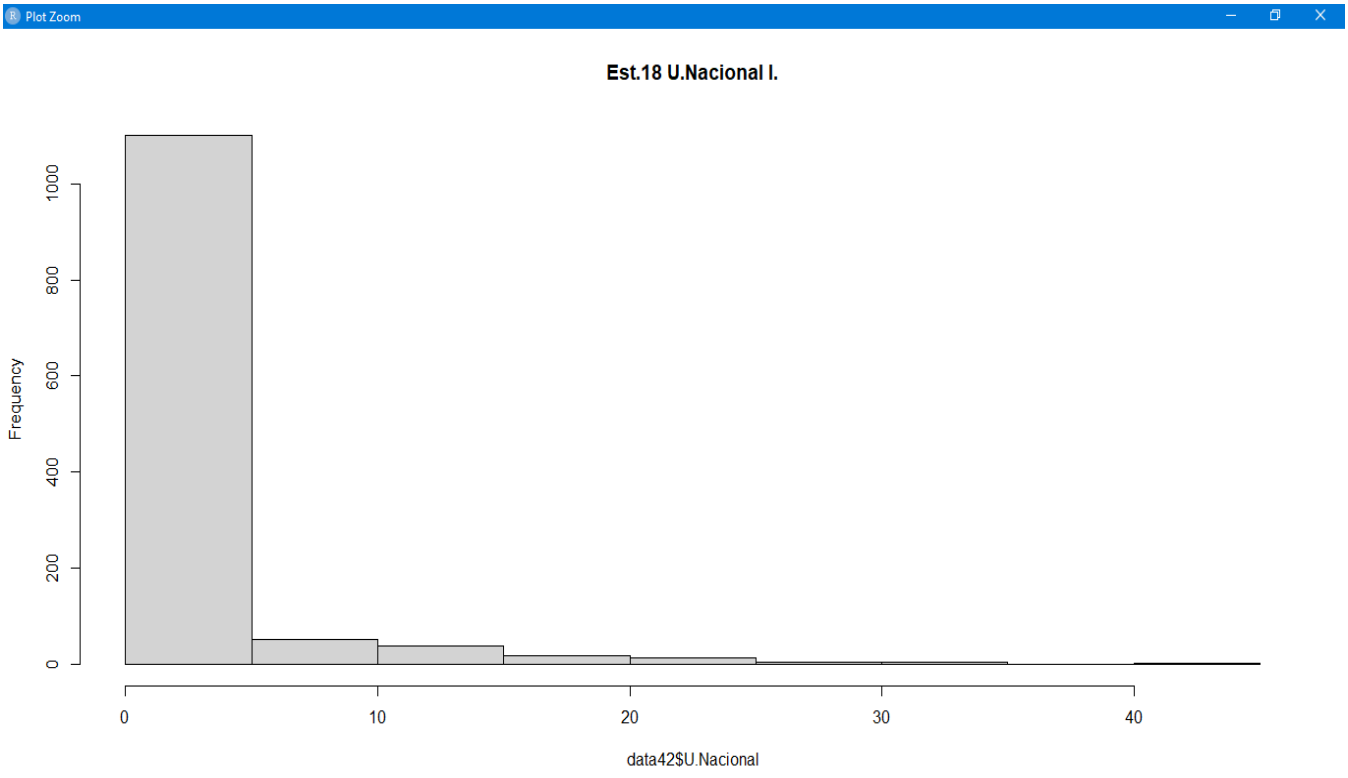
N. 17 Estación S.Francisco



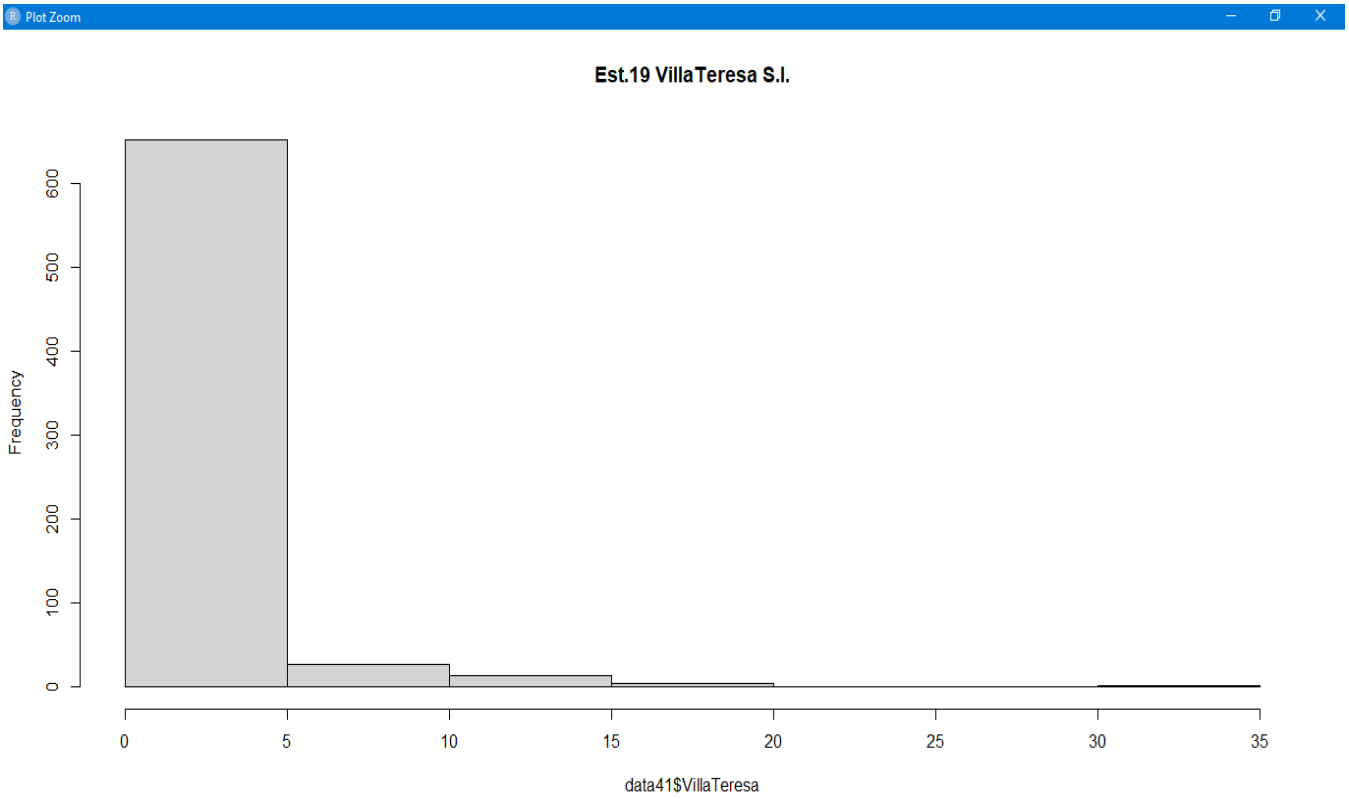


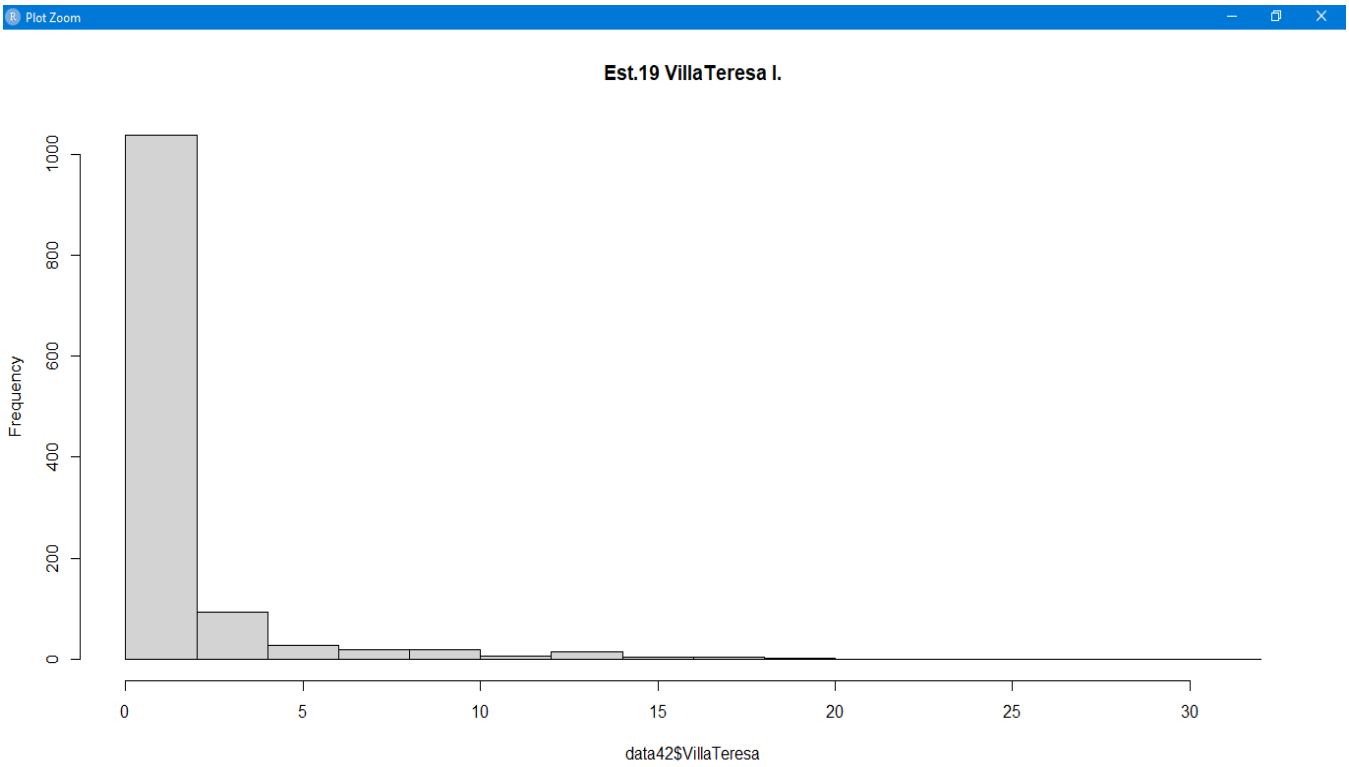
N. 18 Estación U.Nacional





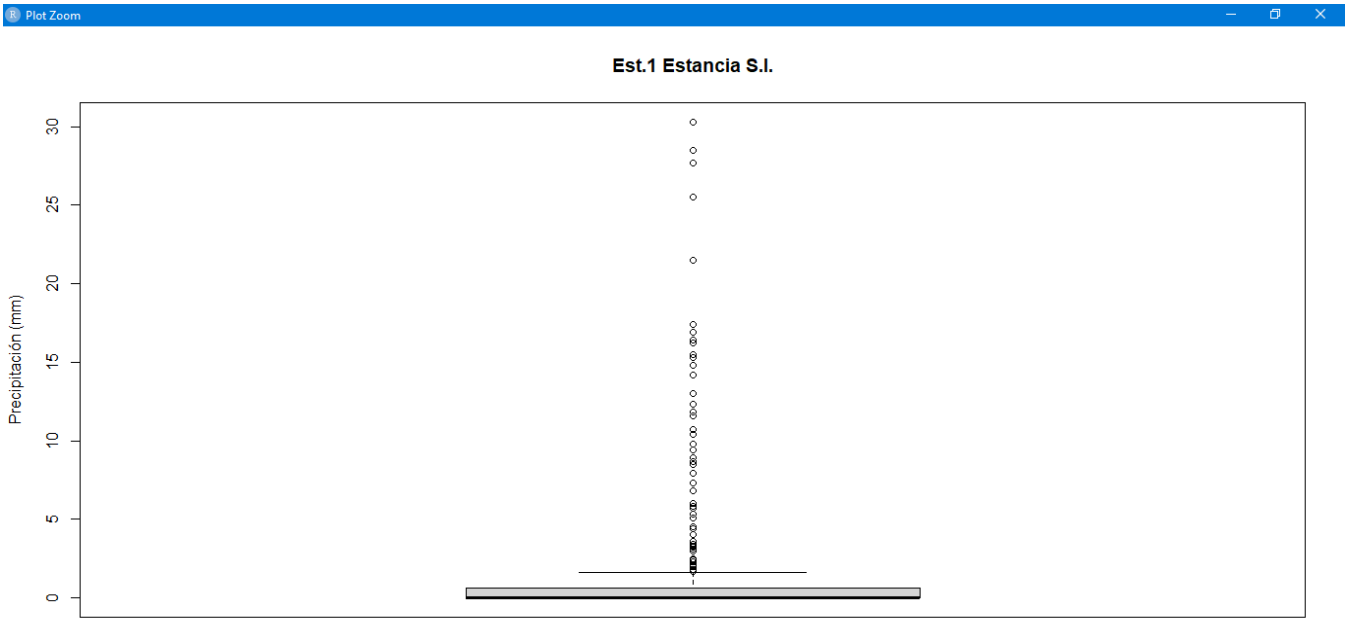
N. 19 Estación VillaTeresa

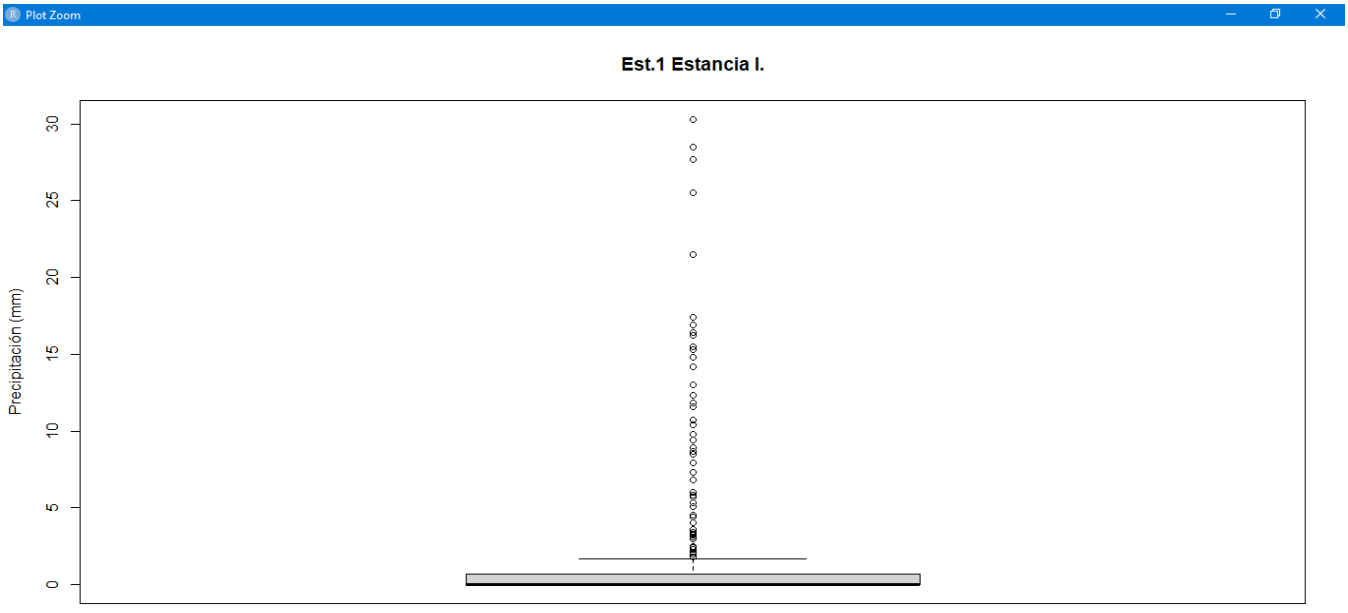




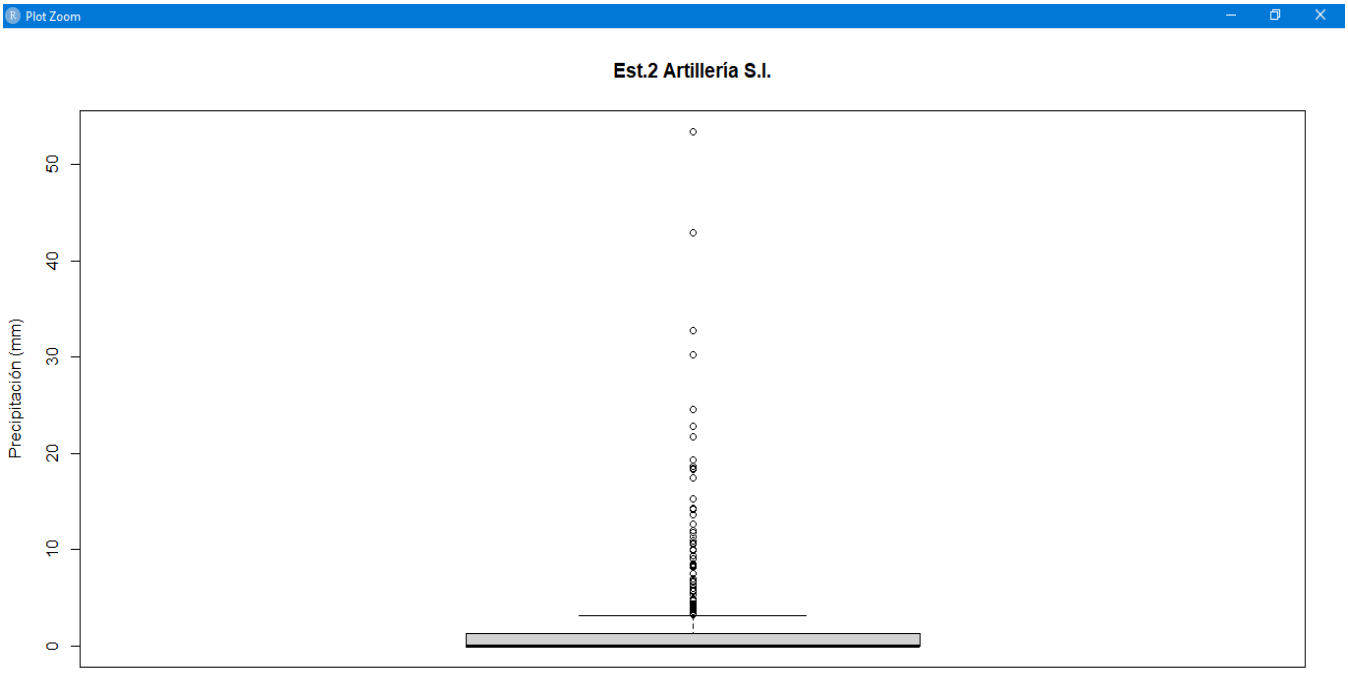
Anexo 3: Boxplots comparativos de las precipitaciones en las 19 estaciones pluviométricas seleccionadas antes (S.I.) y después de ser imputadas (I.).

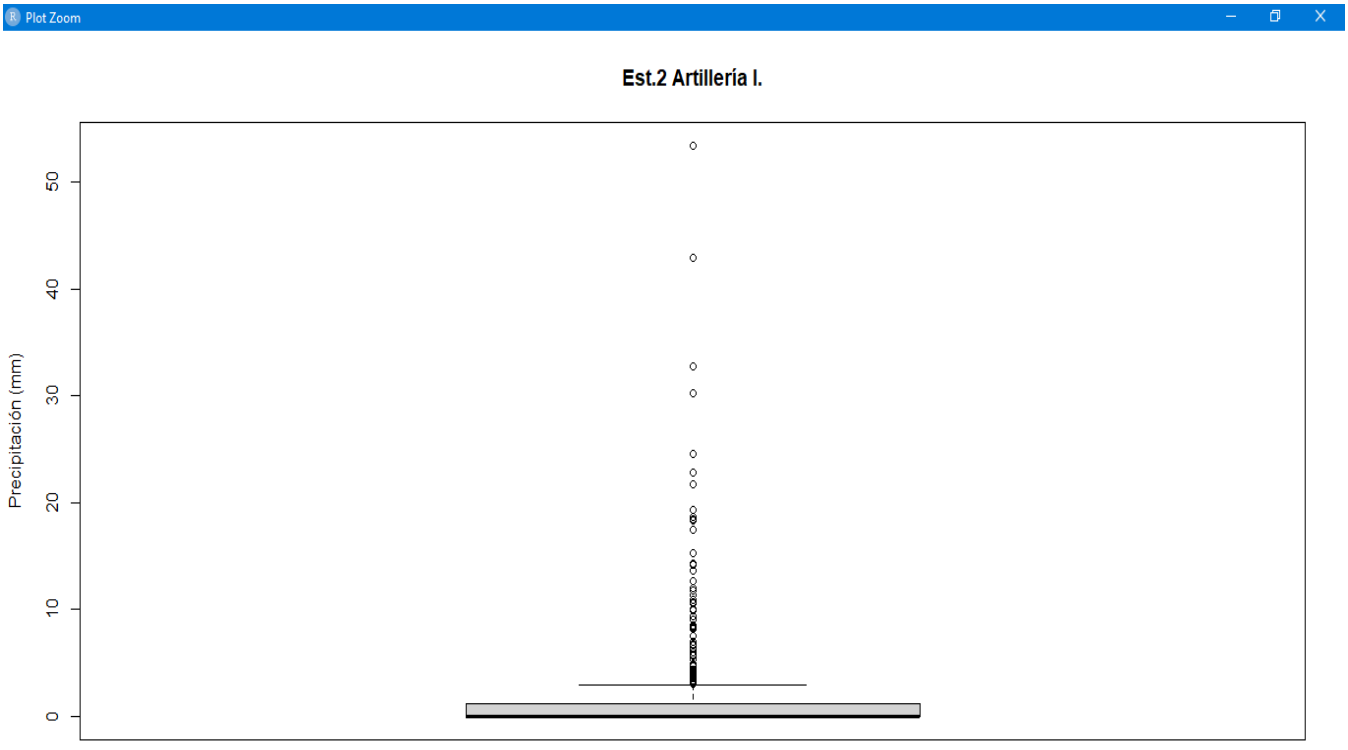
N.1 Estación Estancia



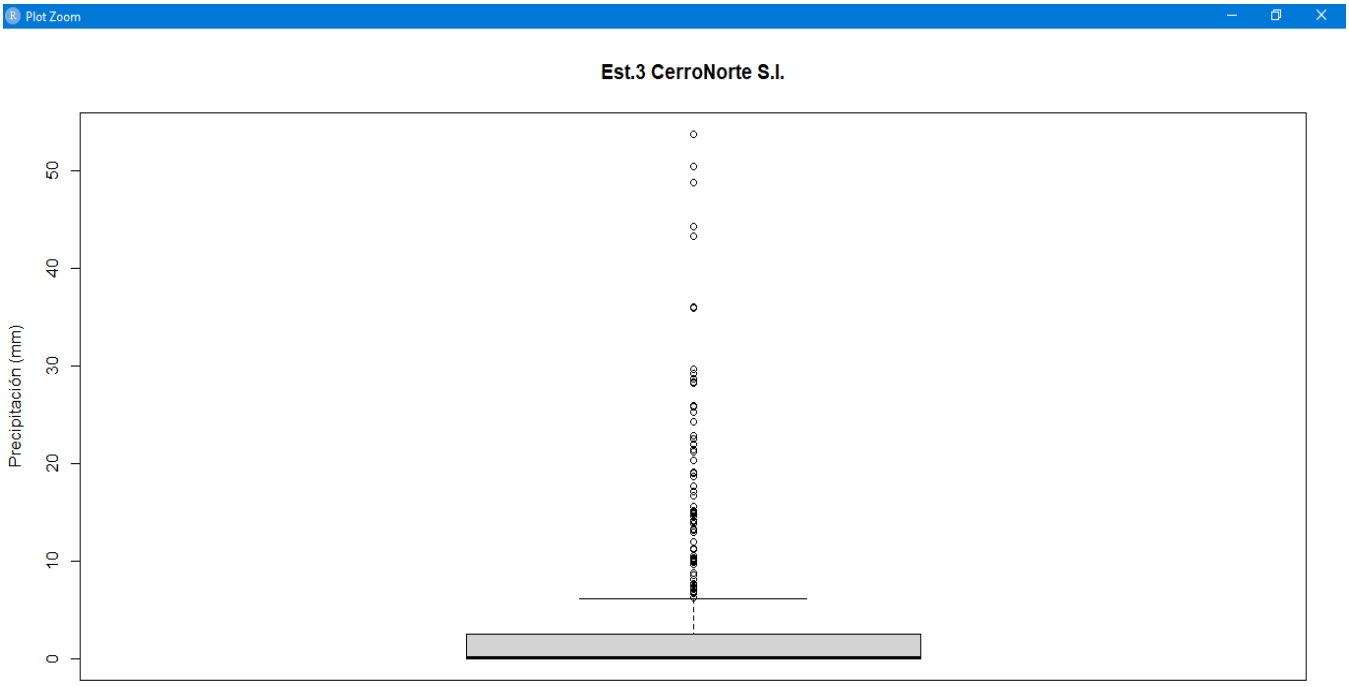


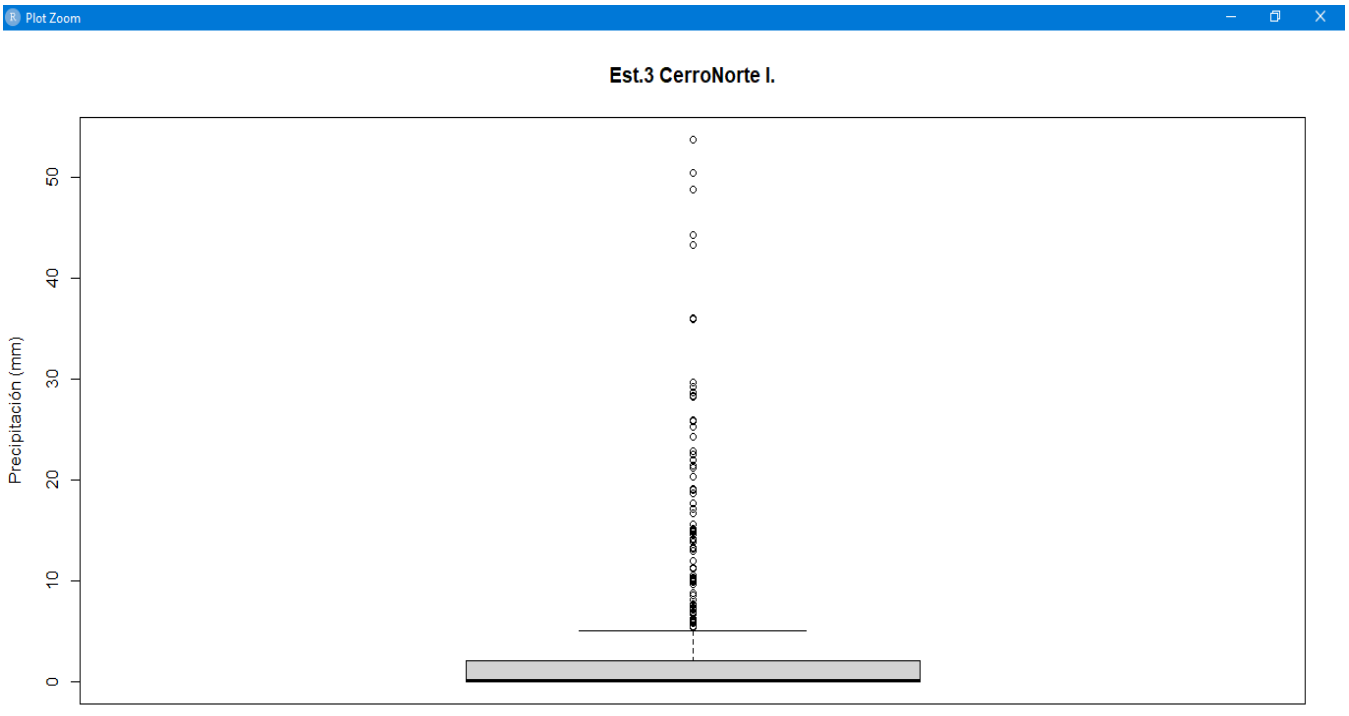
N.2 Estación Artillería



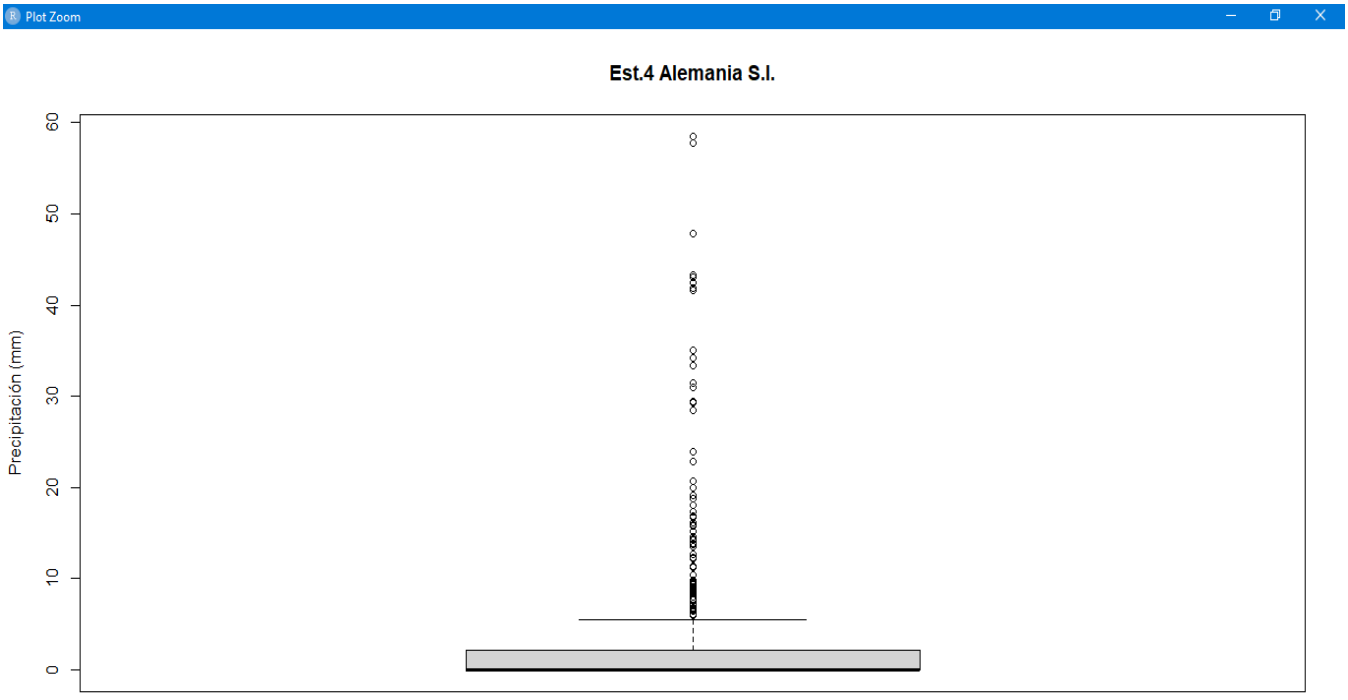


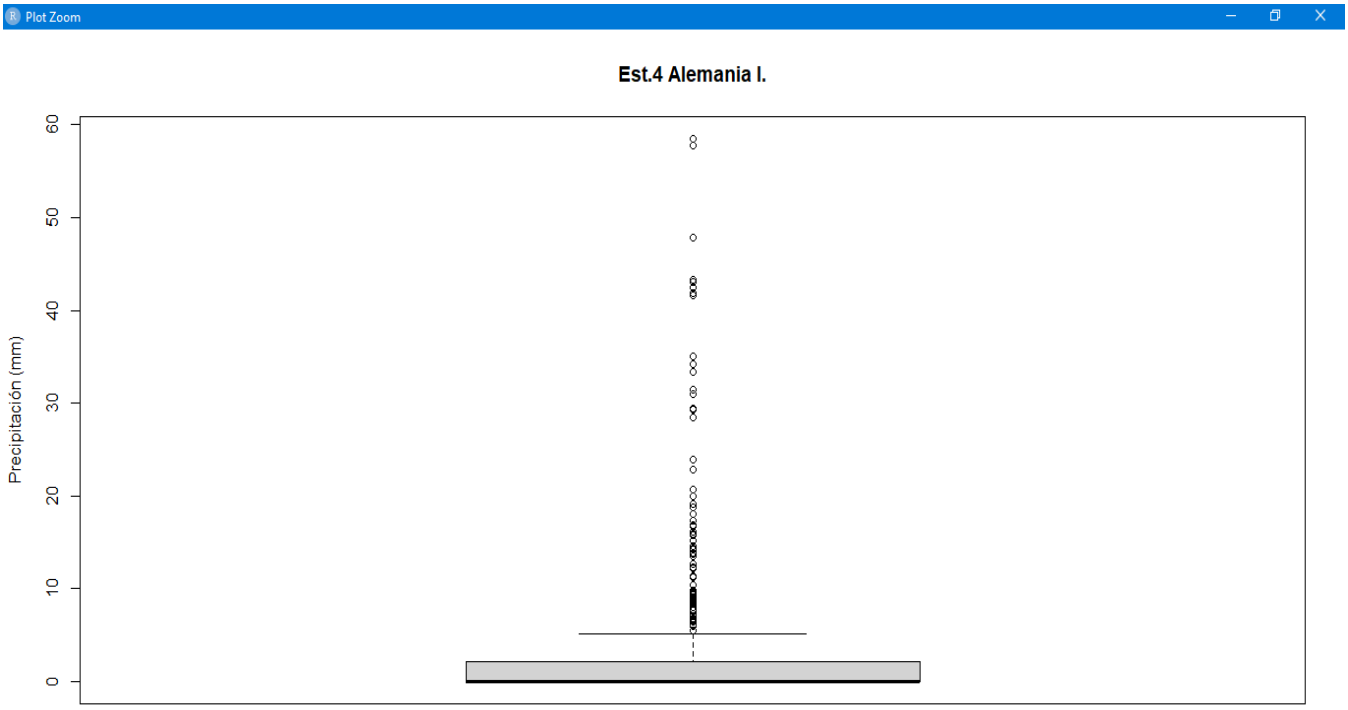
N.3 Estación CerroNorte



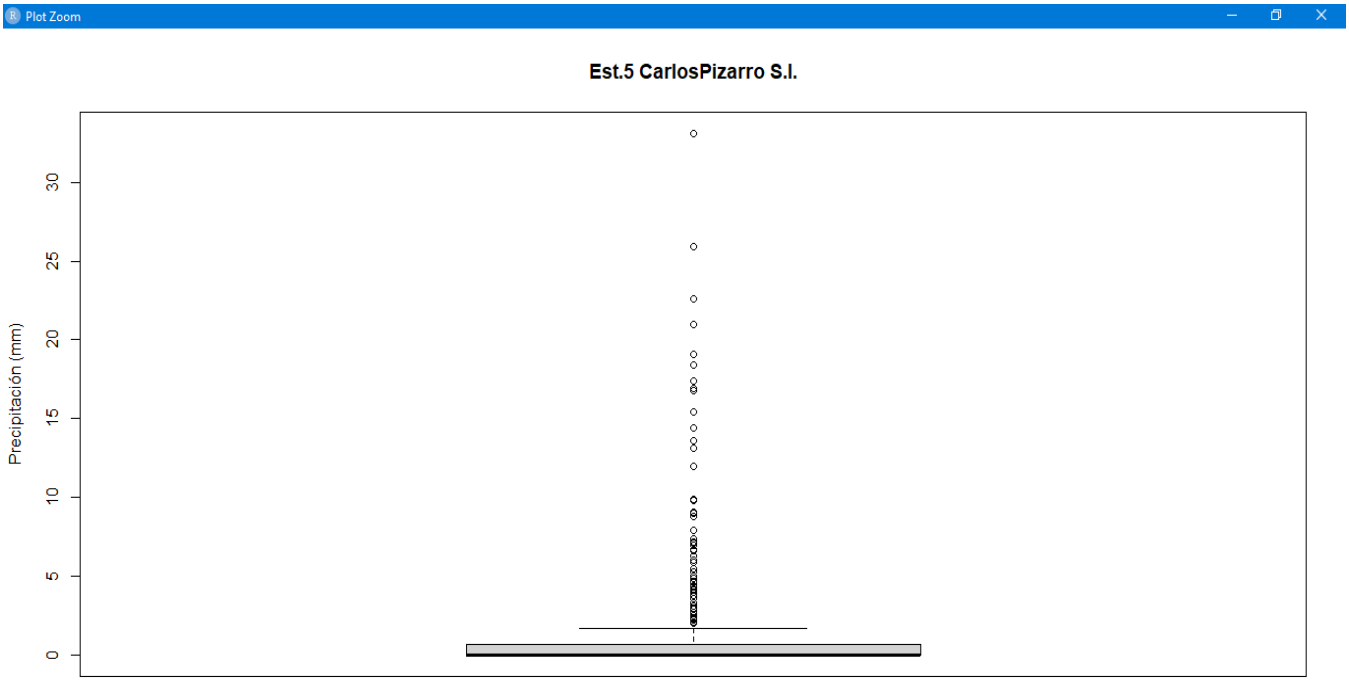


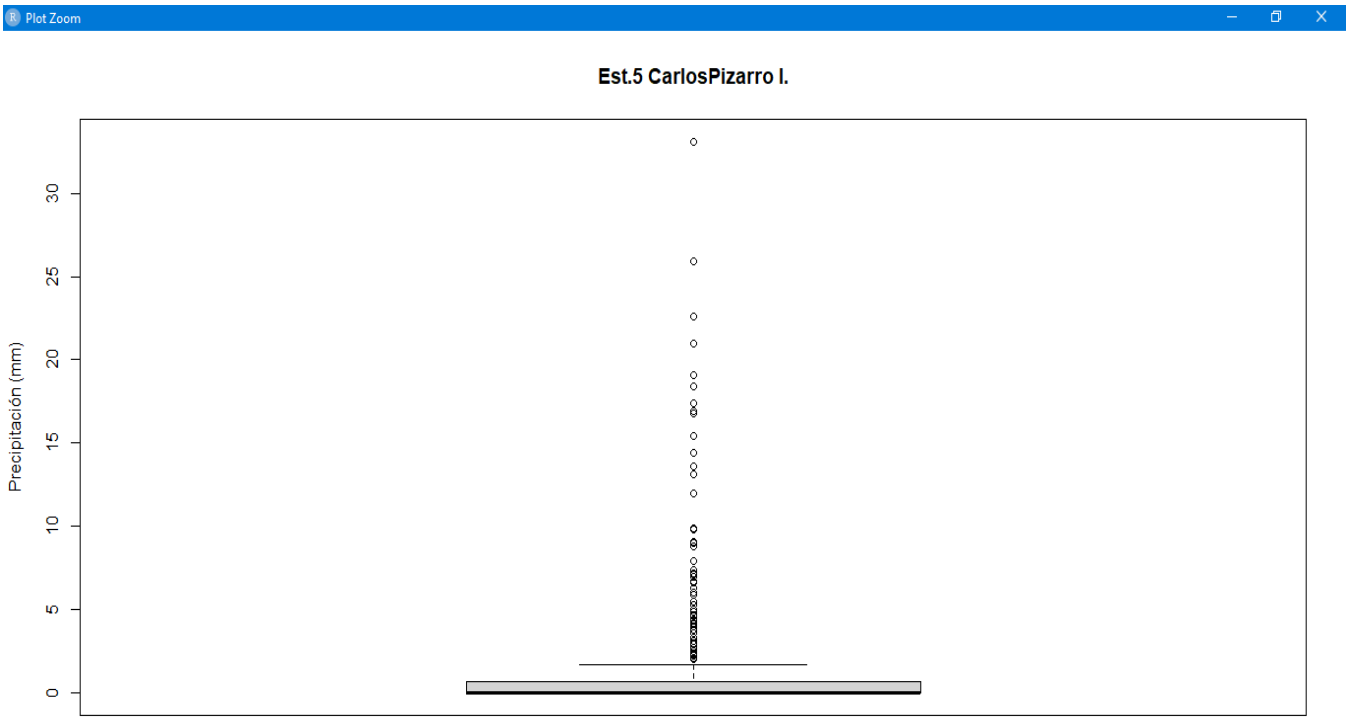
N.4 Estación Alemania



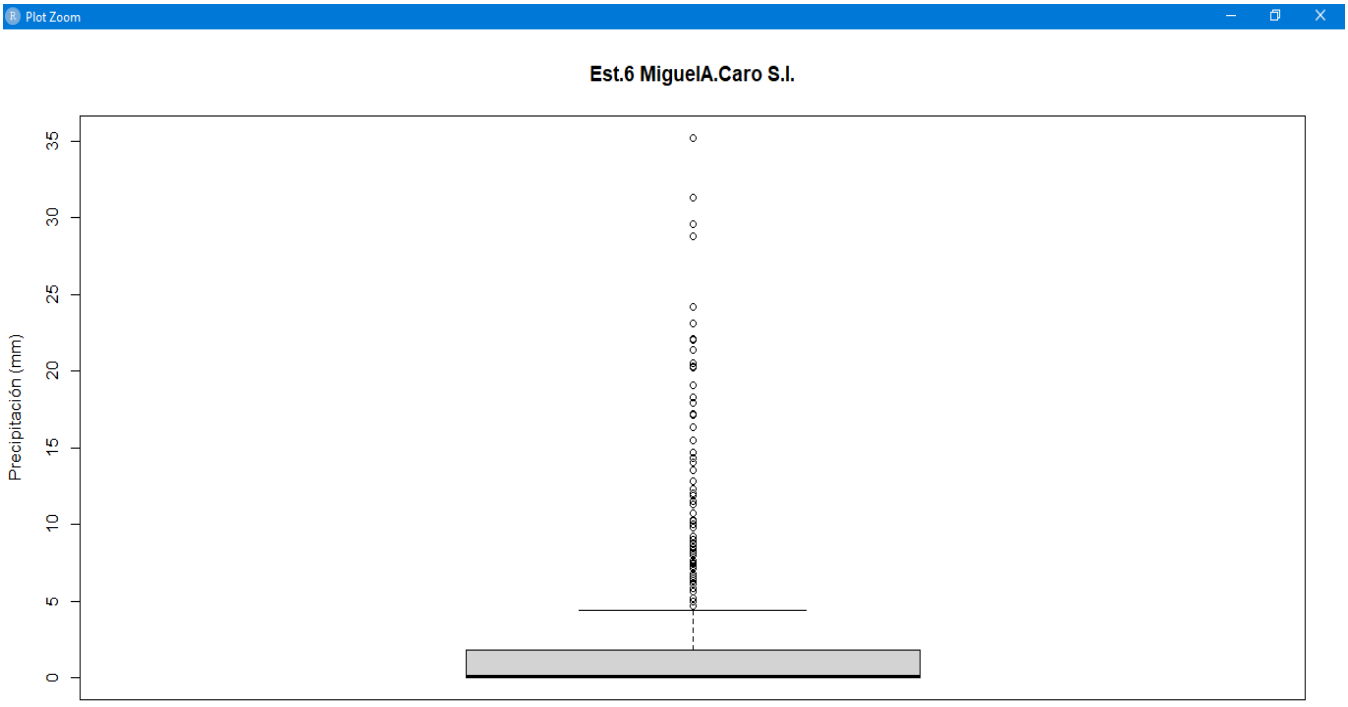


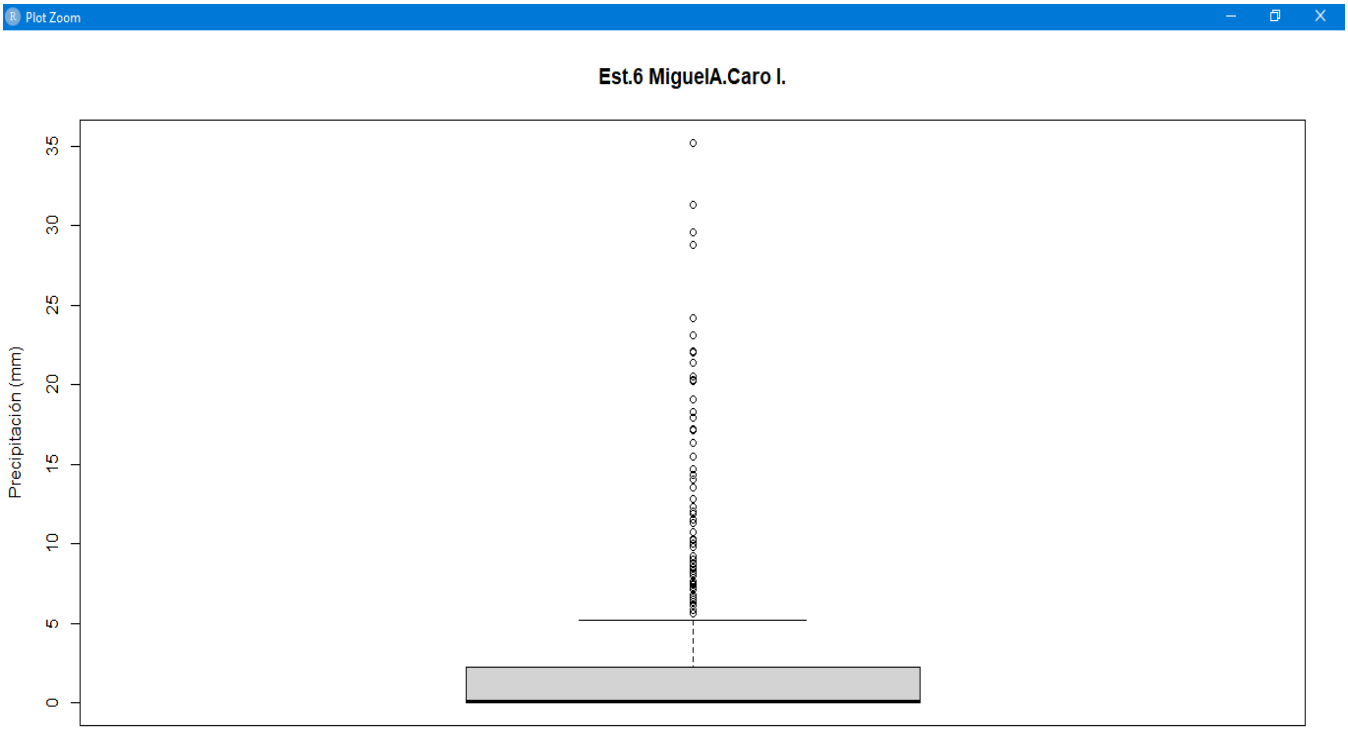
N.5 Estación CarlosPizarro



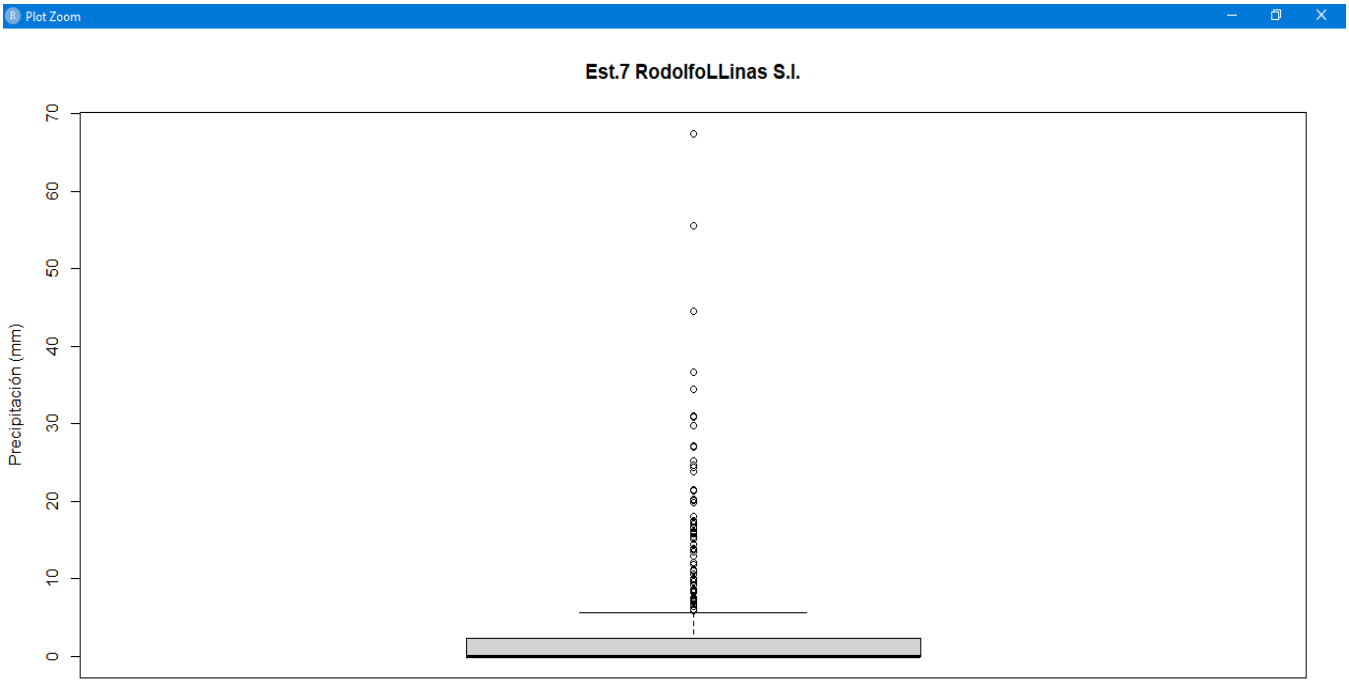


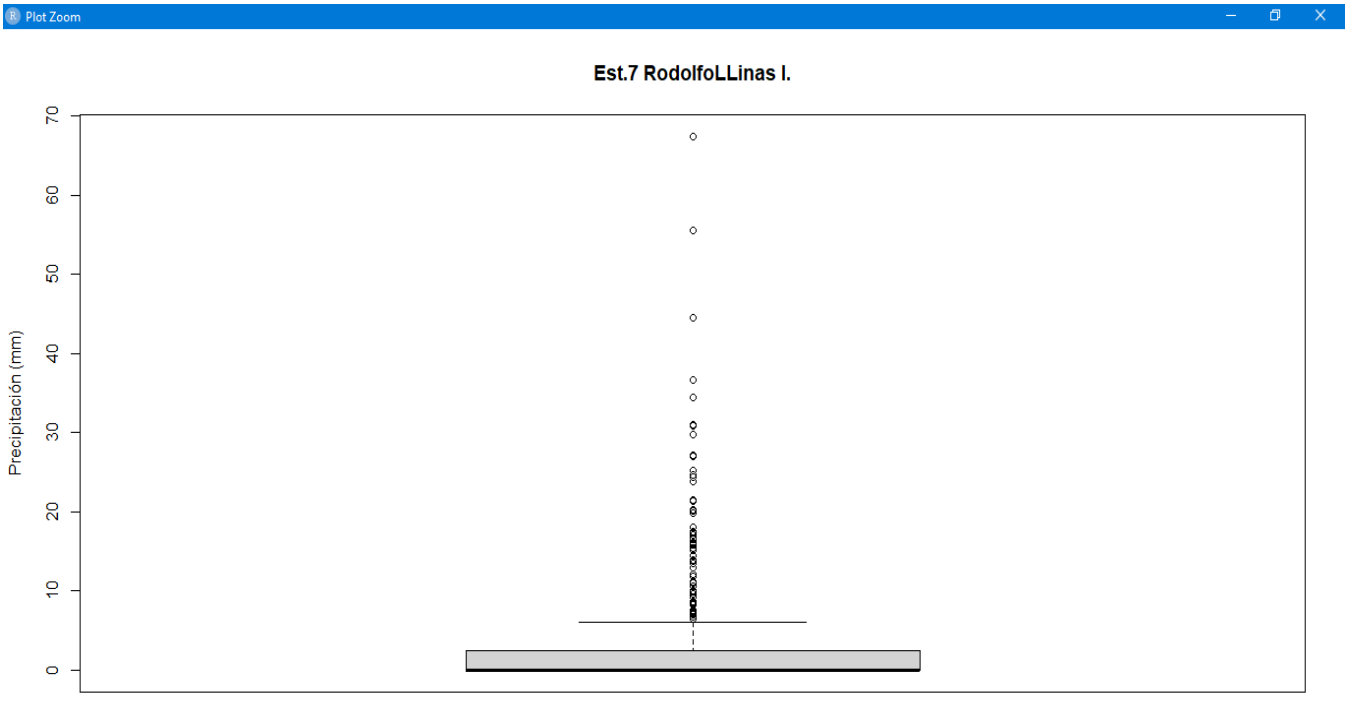
N.6 Estación MiguelACaro



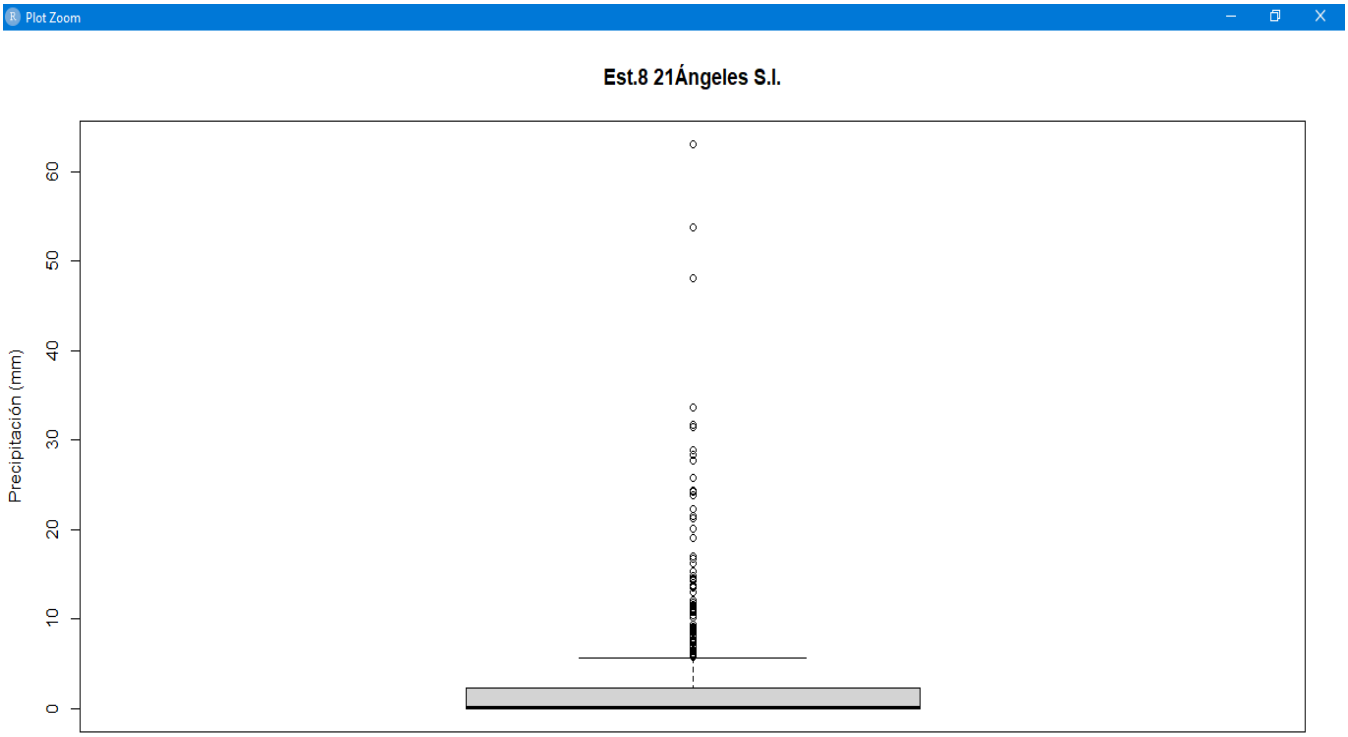


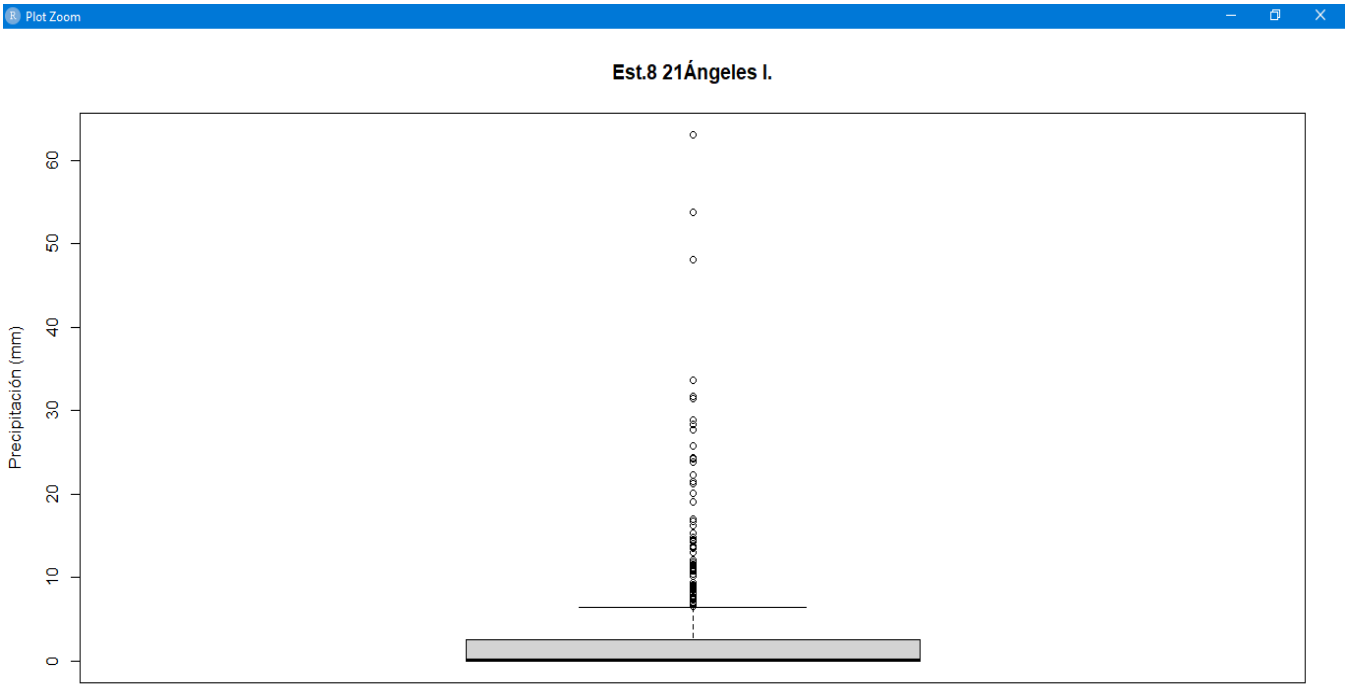
N. 7 Estación RodolfoLLinas



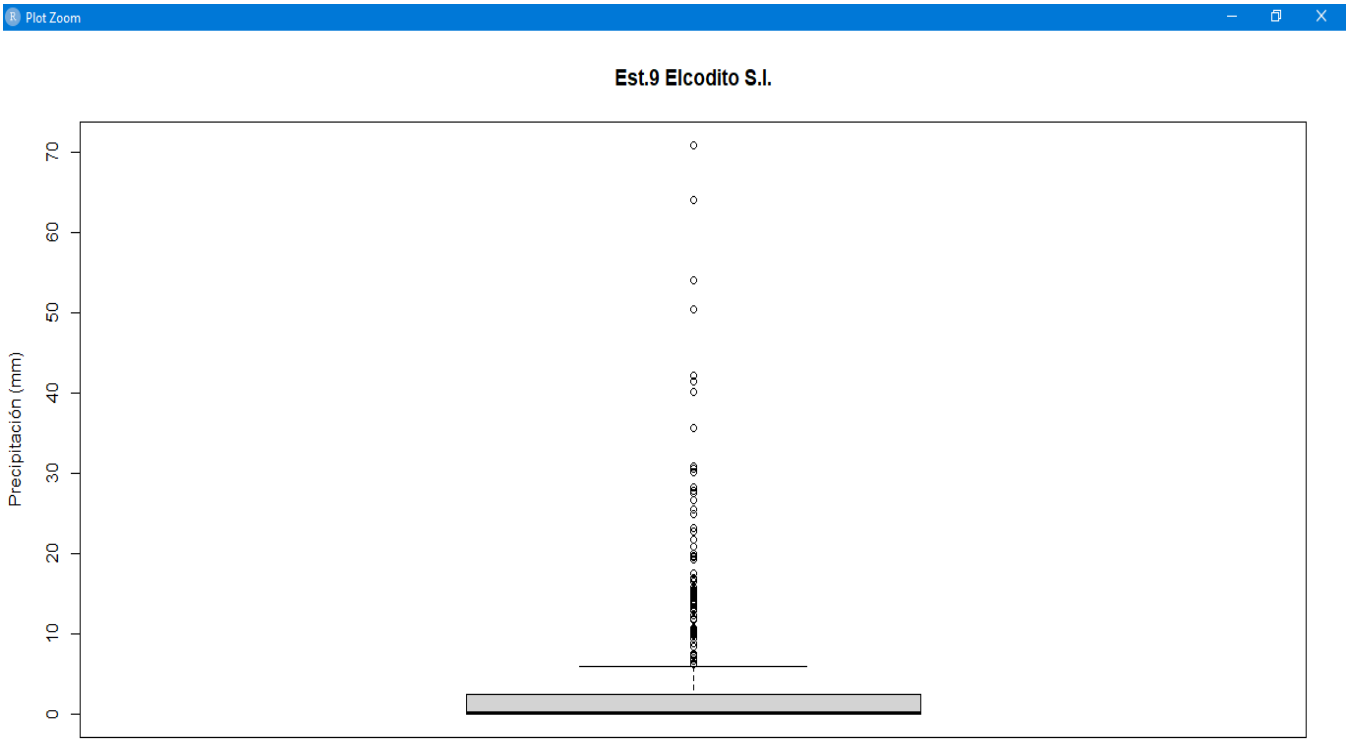


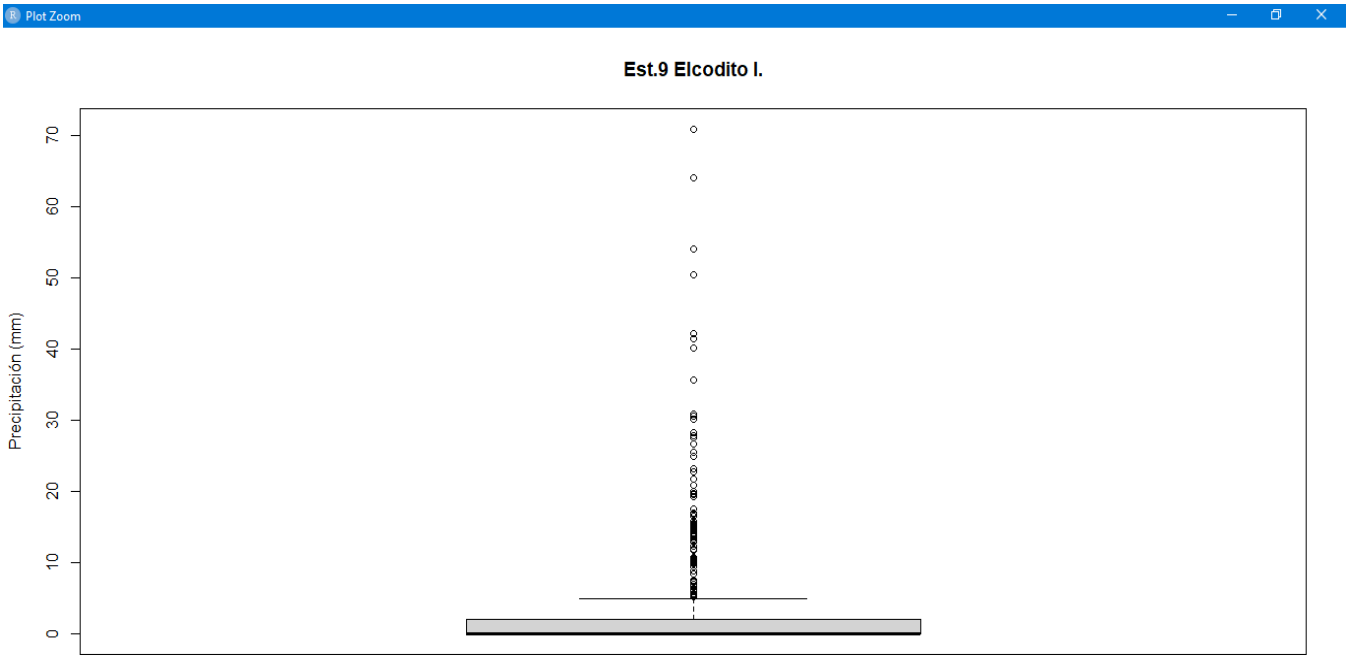
N. 8 Estación 21Ángeles



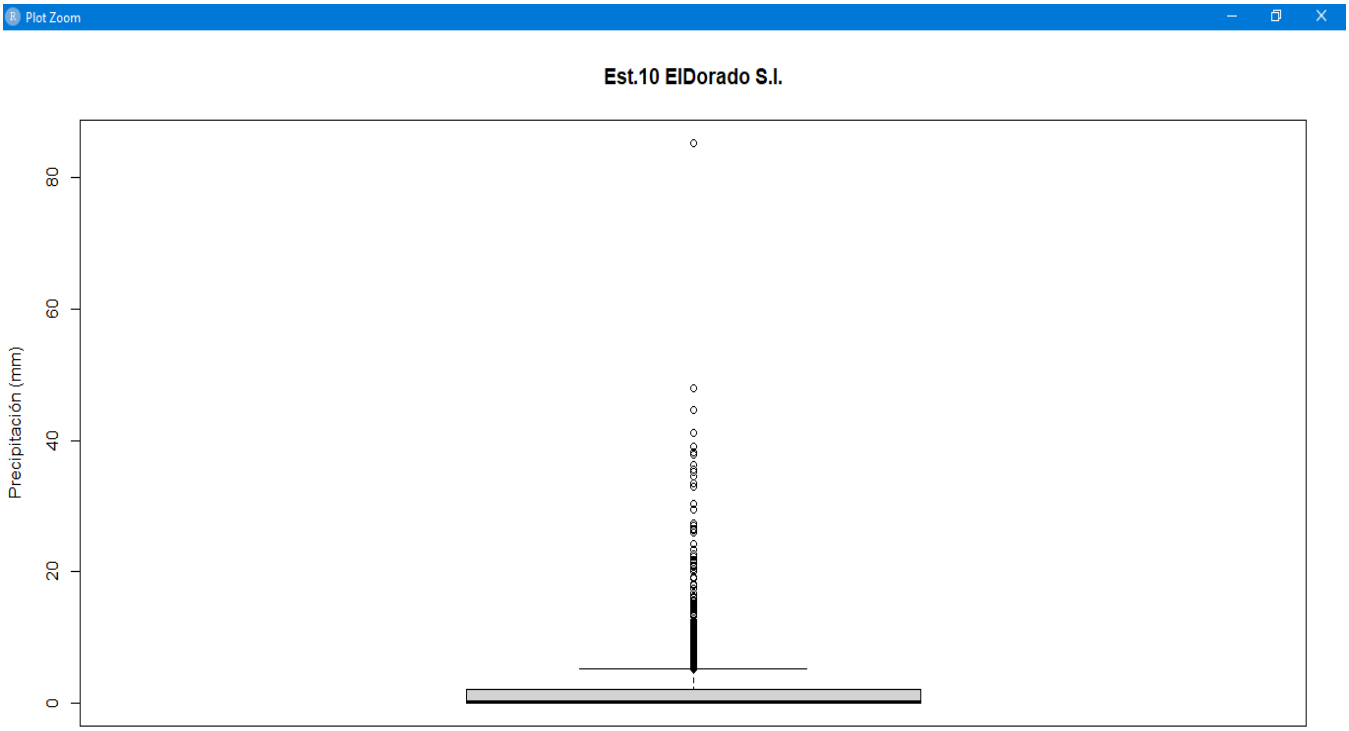


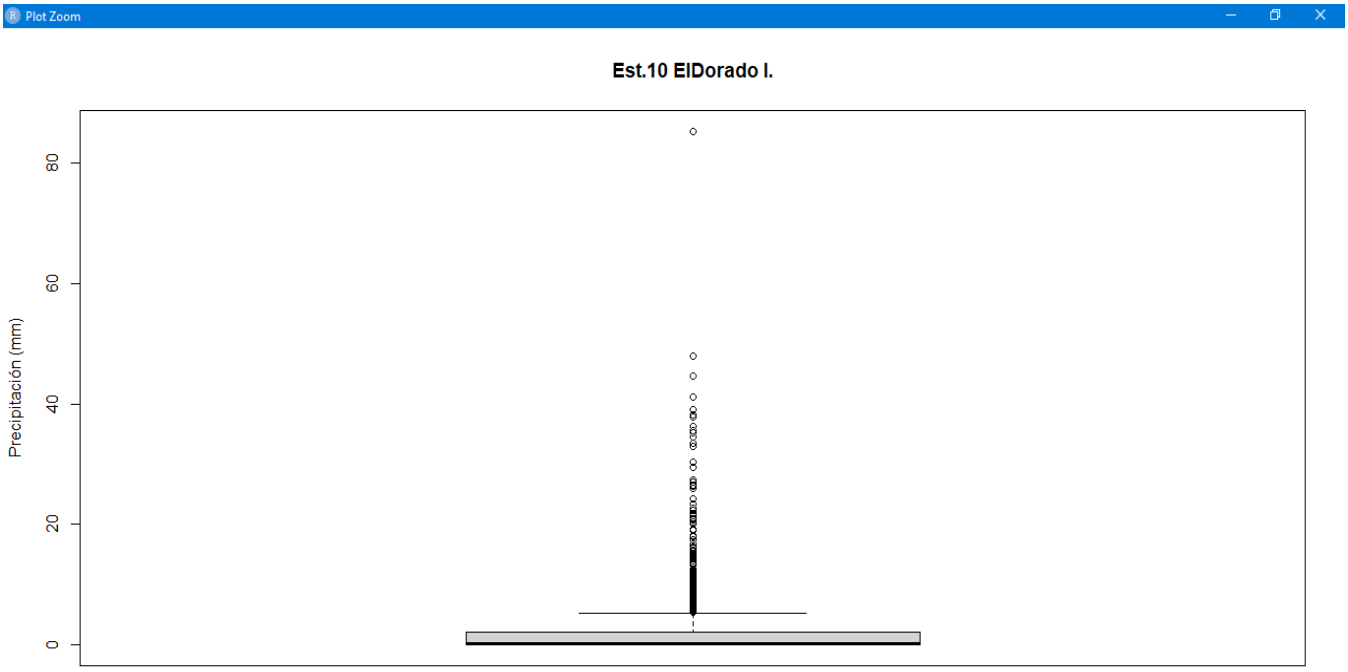
N. 9 Estación ElCodito



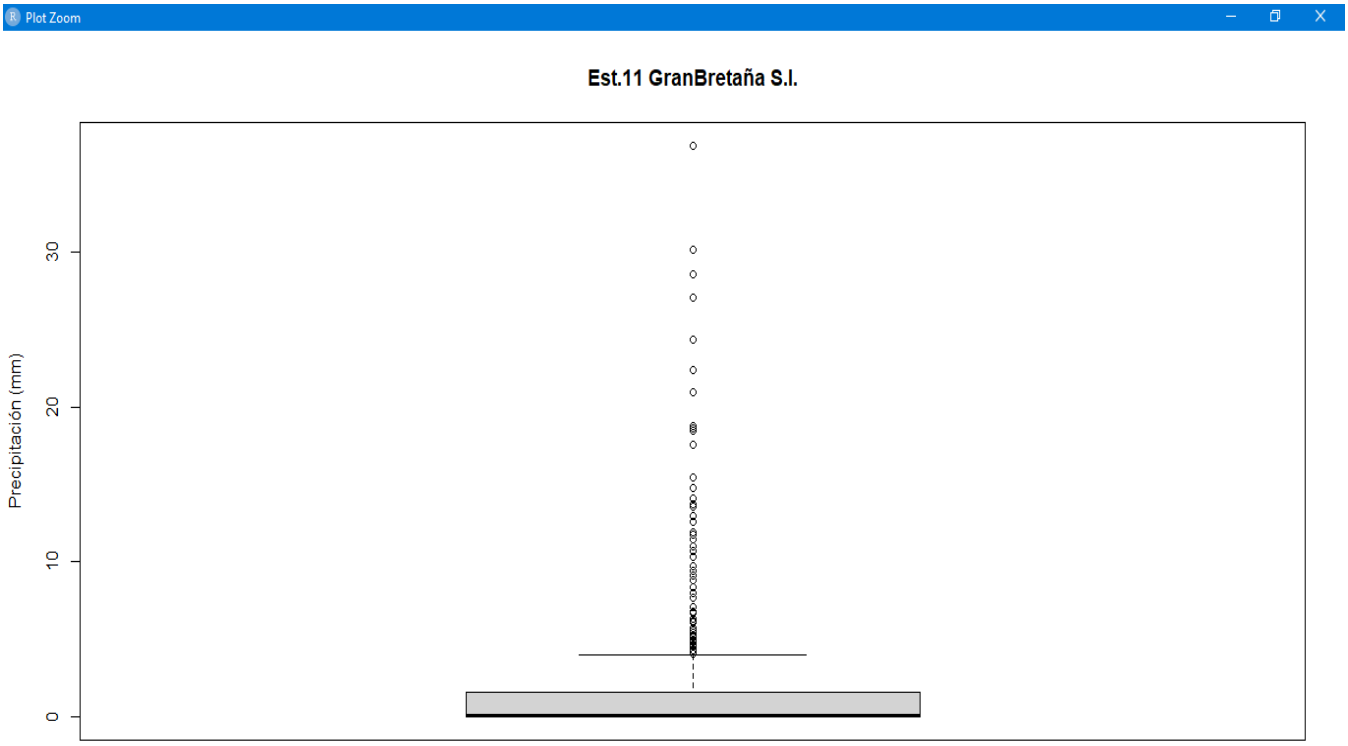


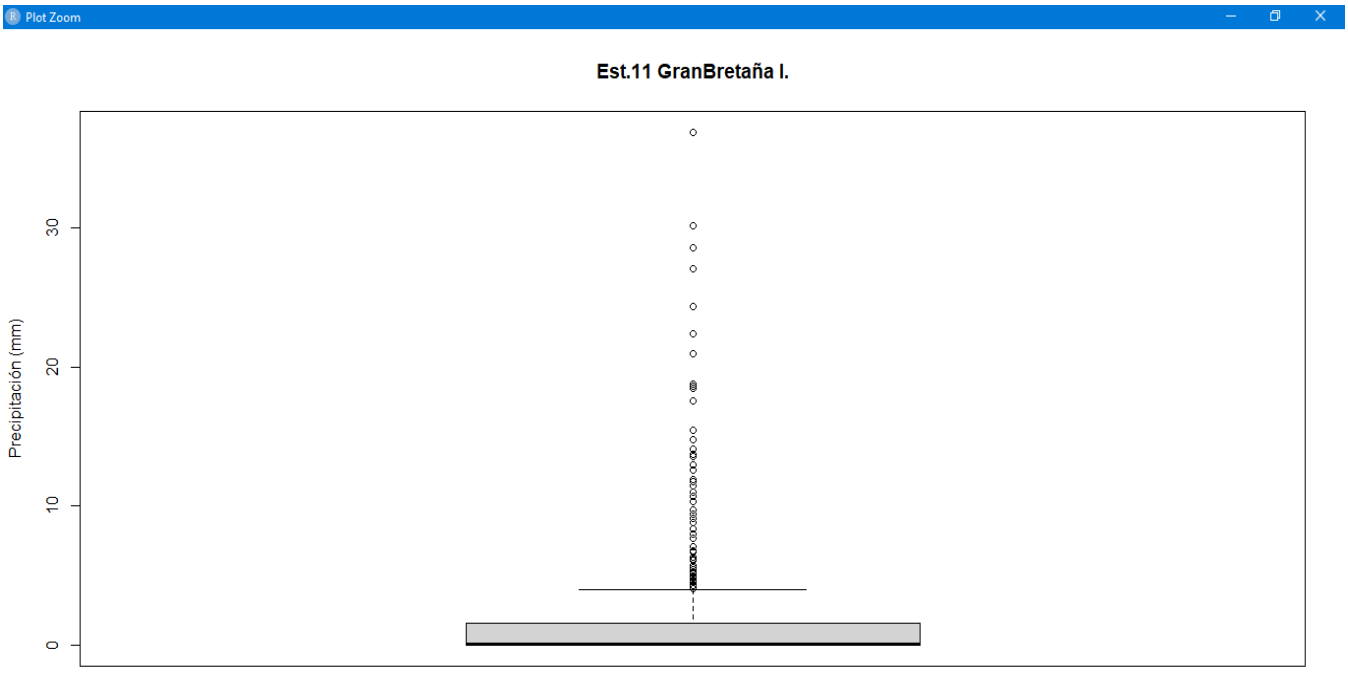
N. 10 Estación ElDorado



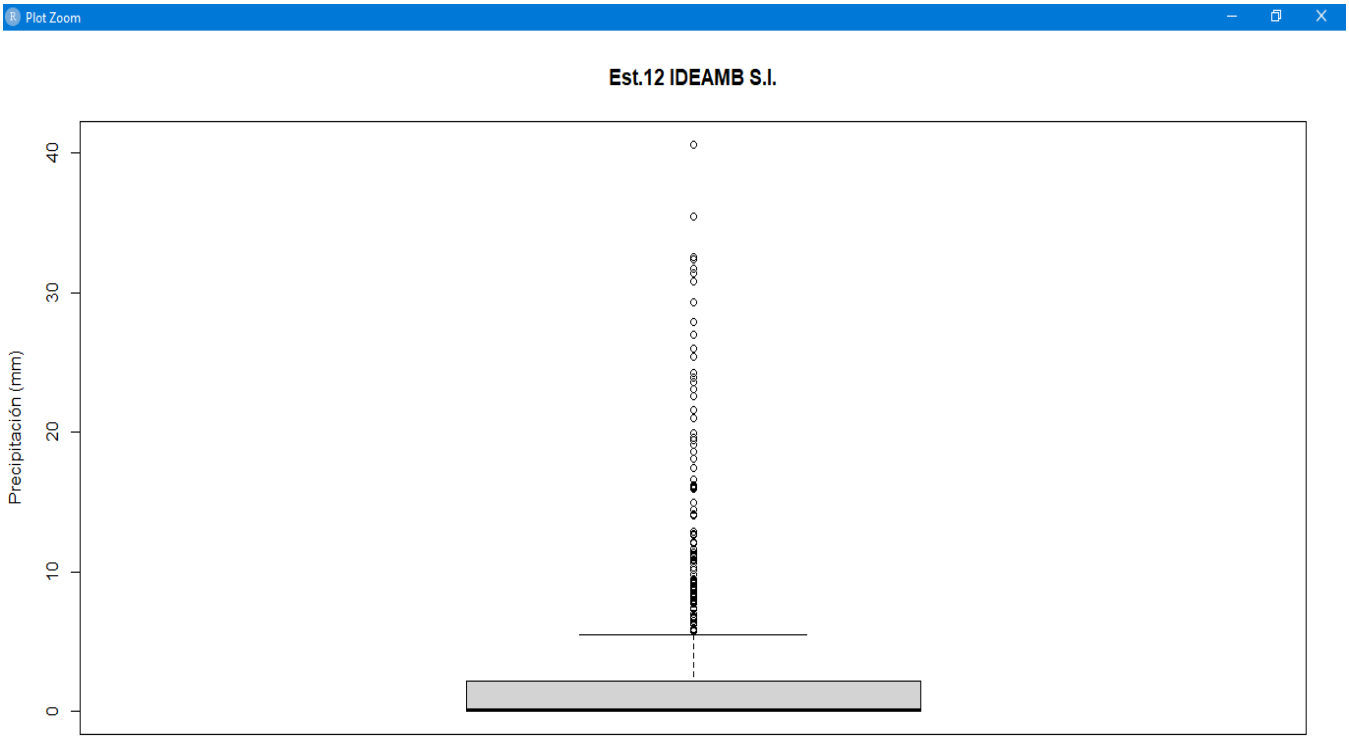


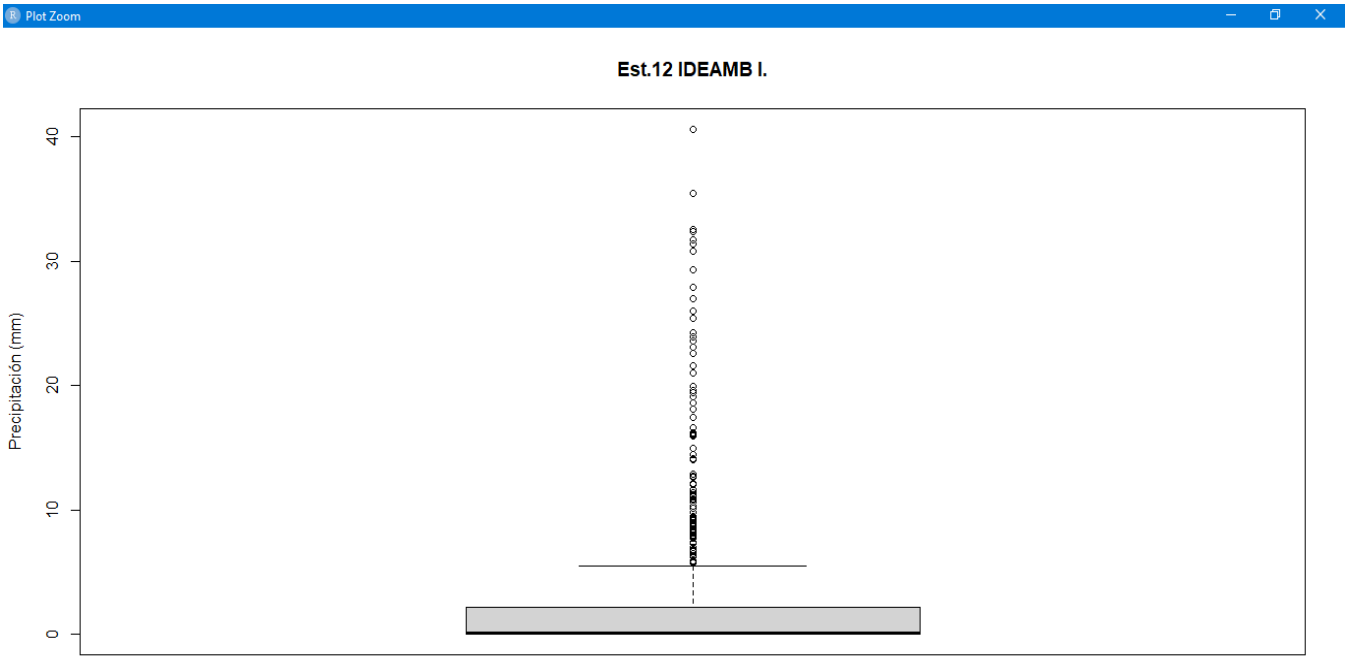
N. 11 Estación Bretaña



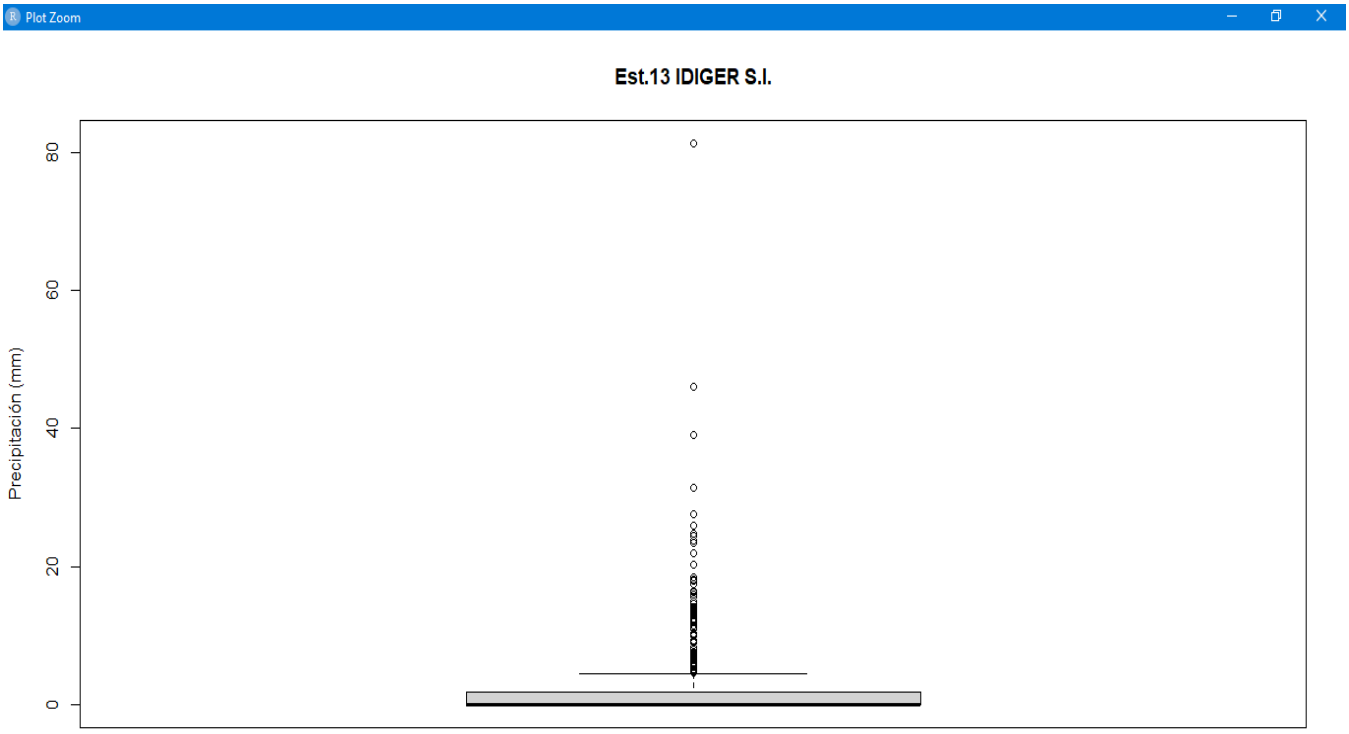


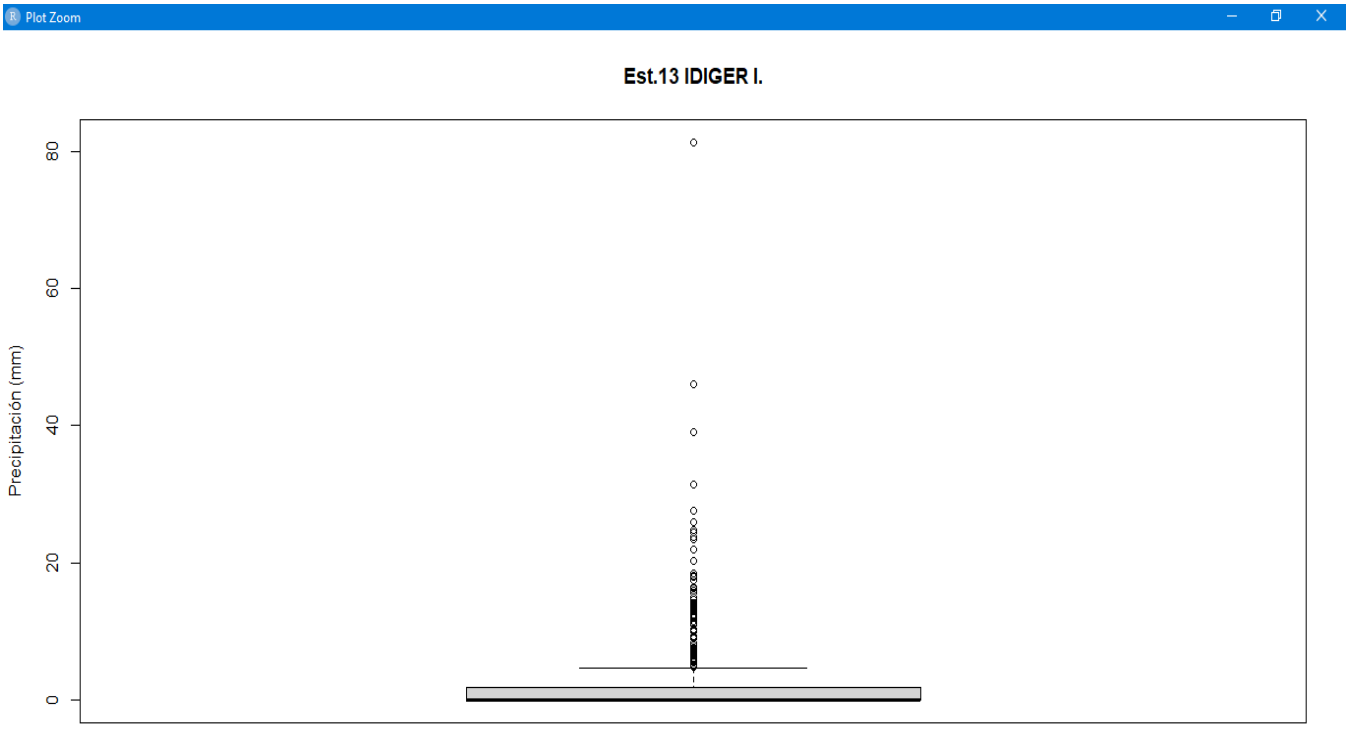
N. 12 Estación IDEAMB



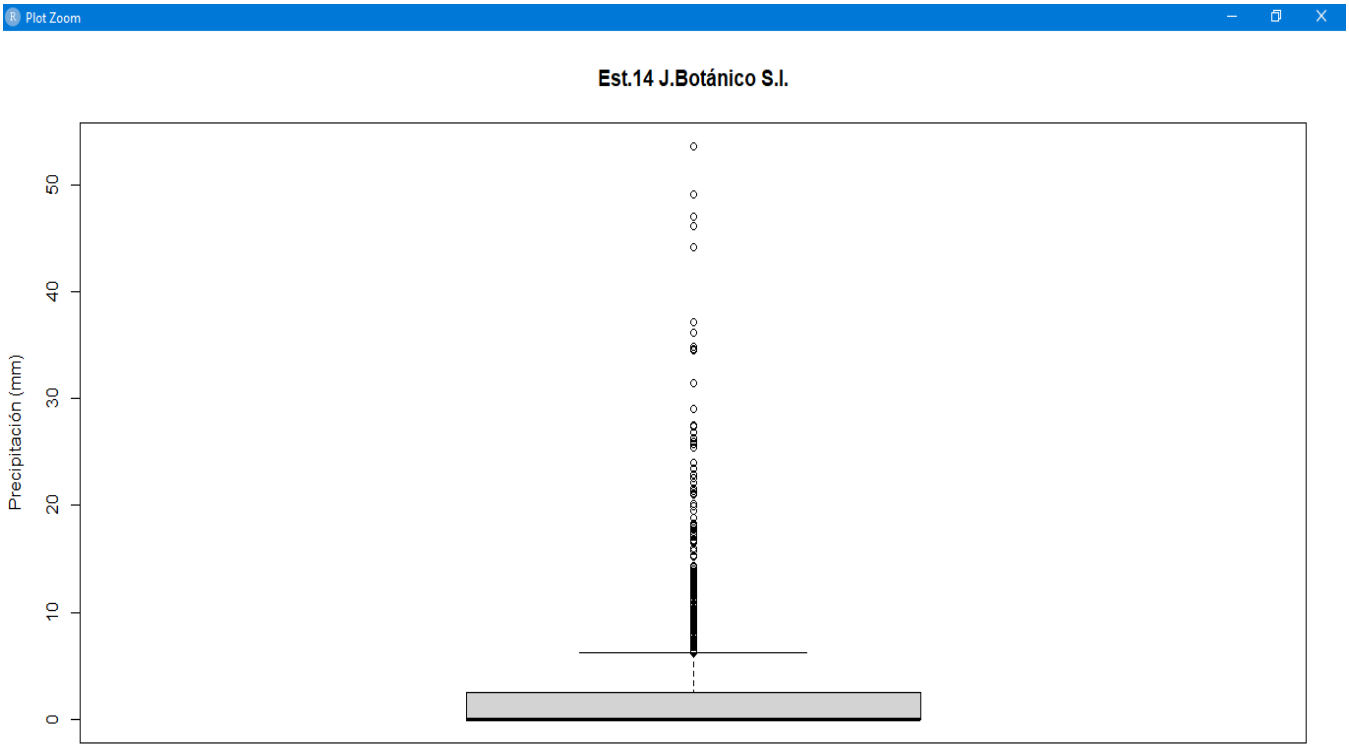


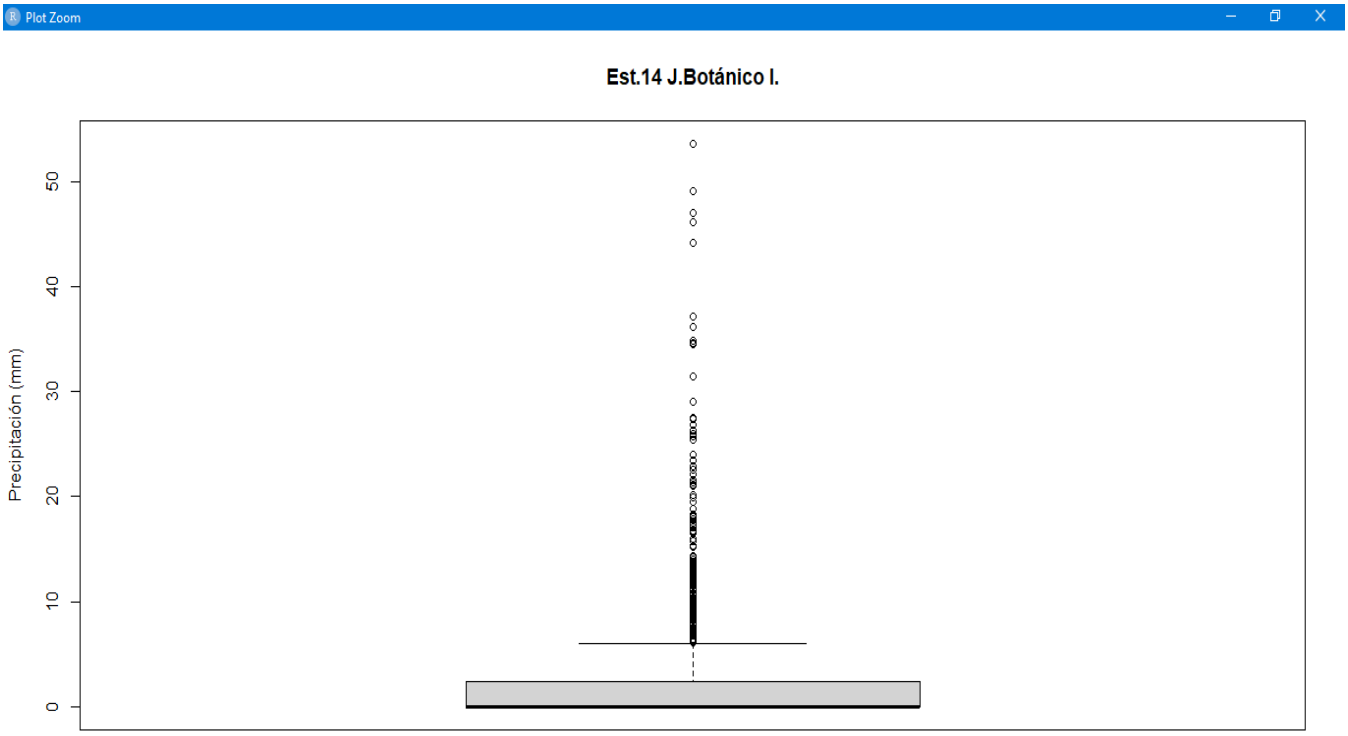
N. 13 Estación IDIGER



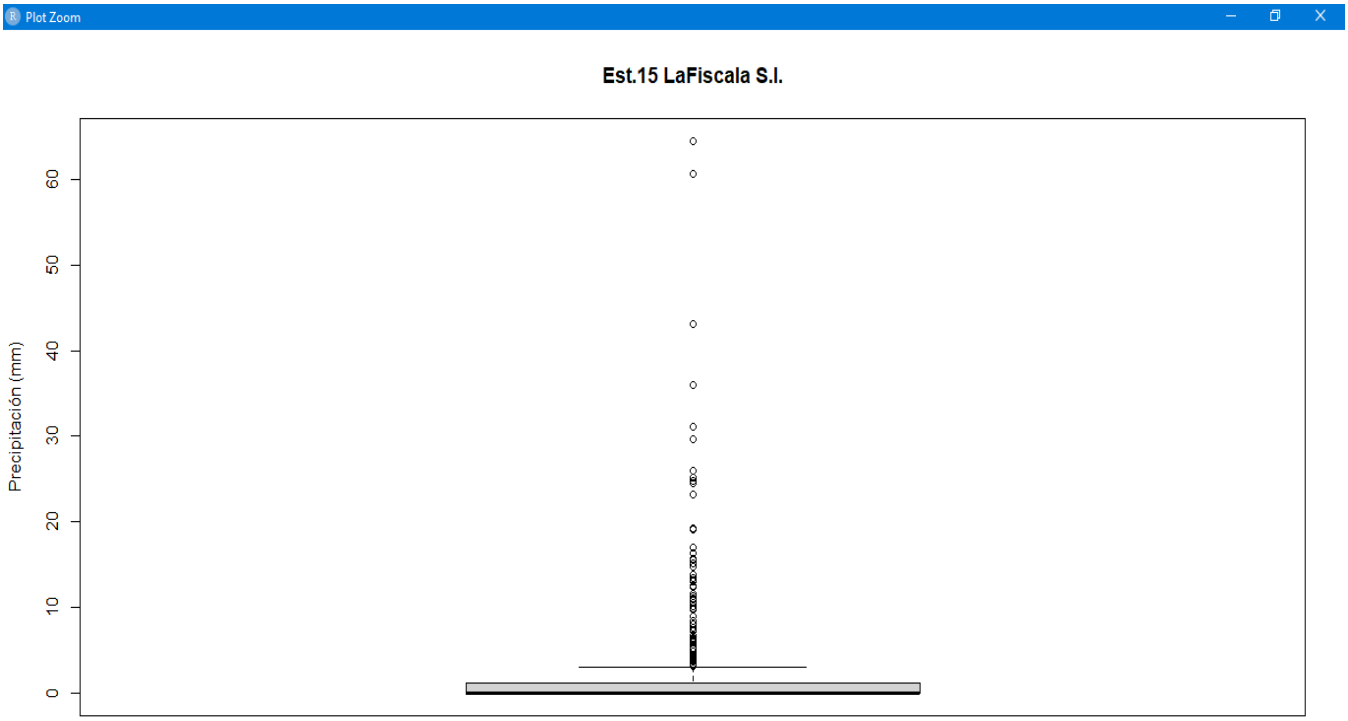


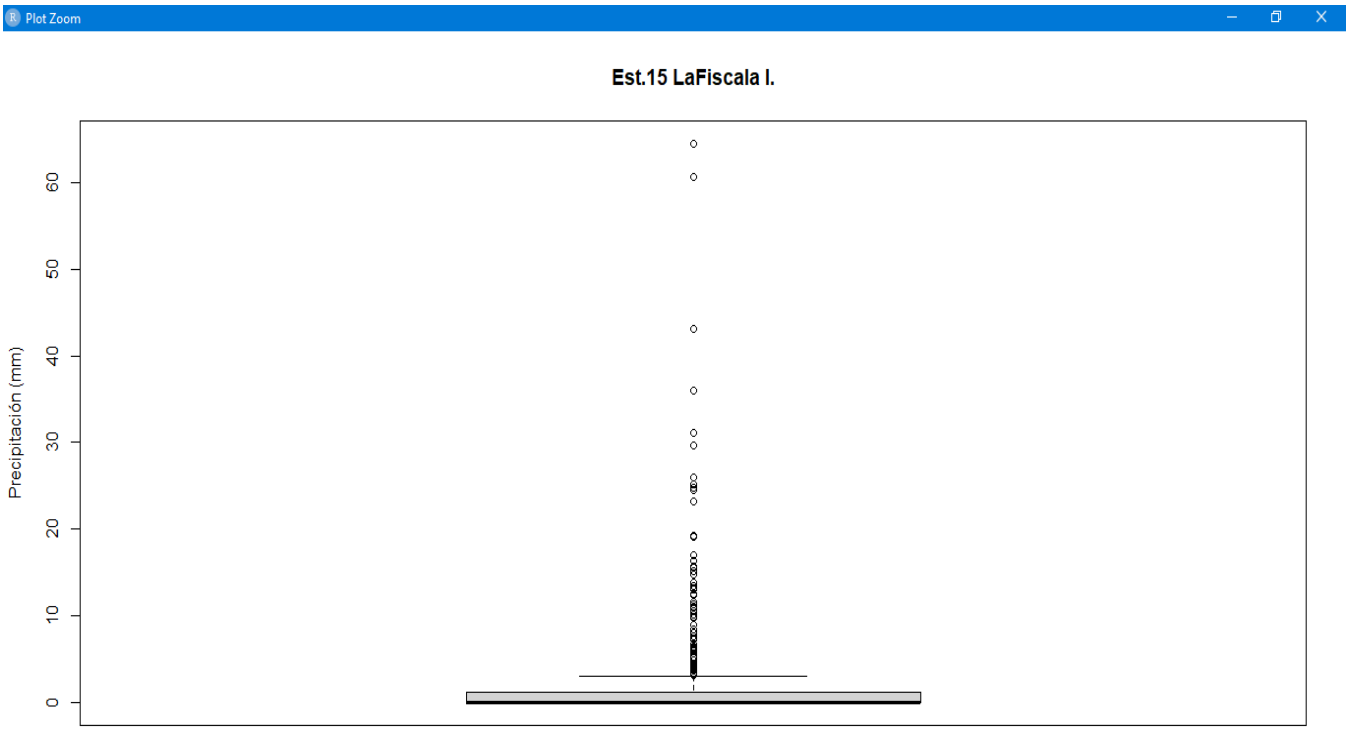
N. 14 Estación JBotánico



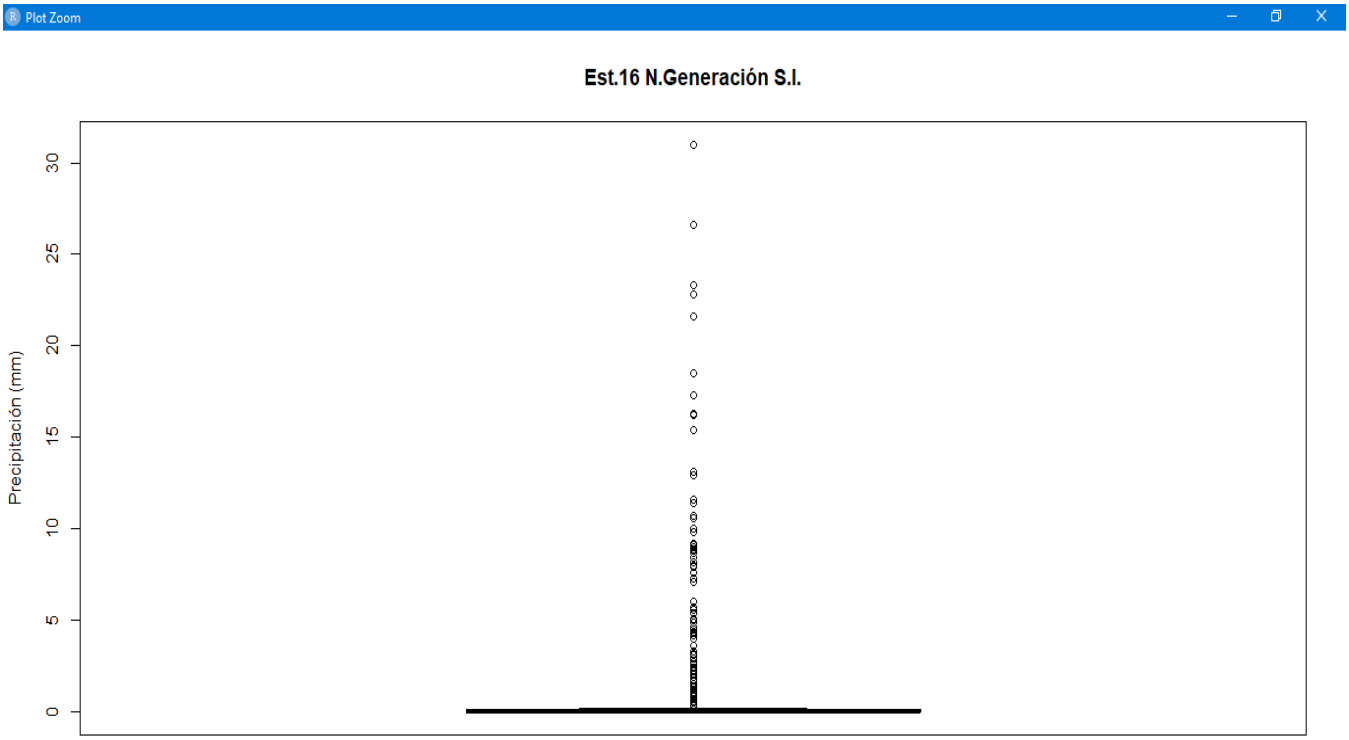


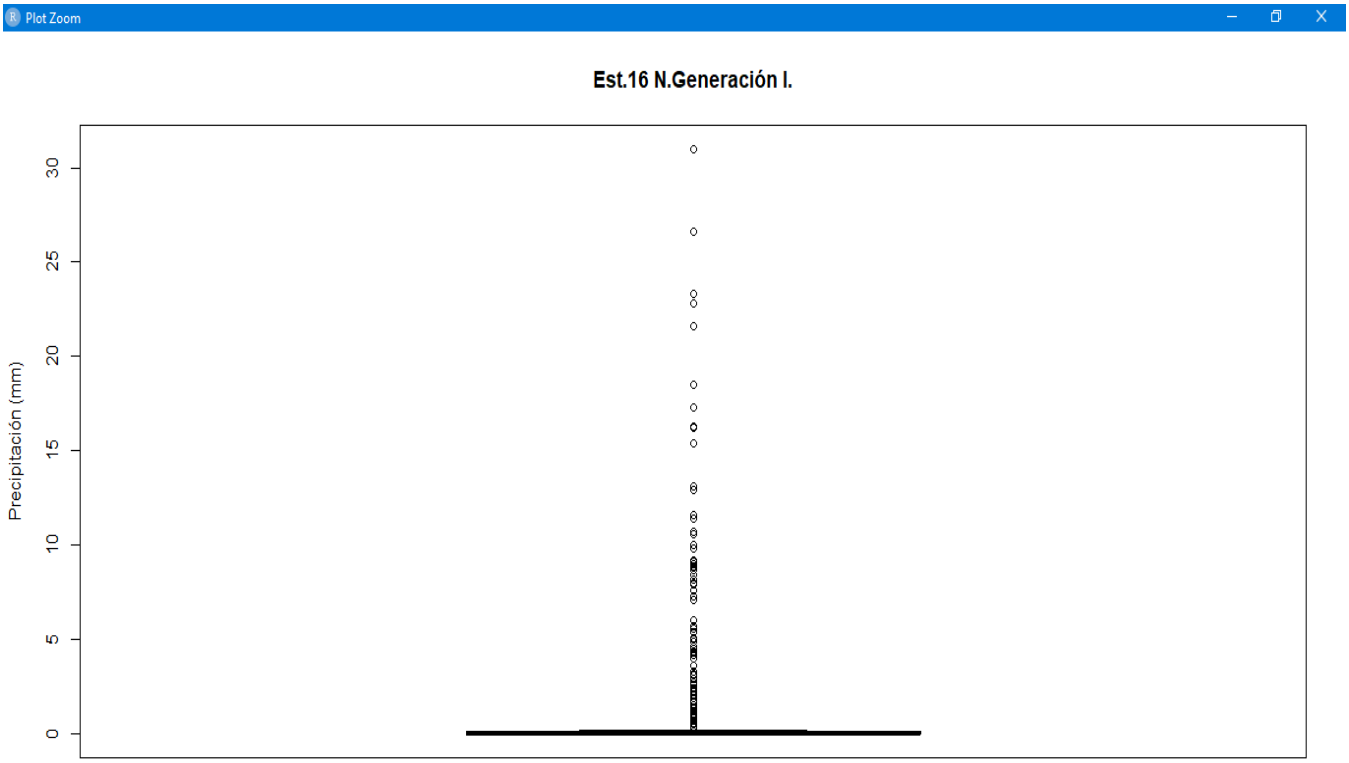
N. 15 Estación LaFiscala



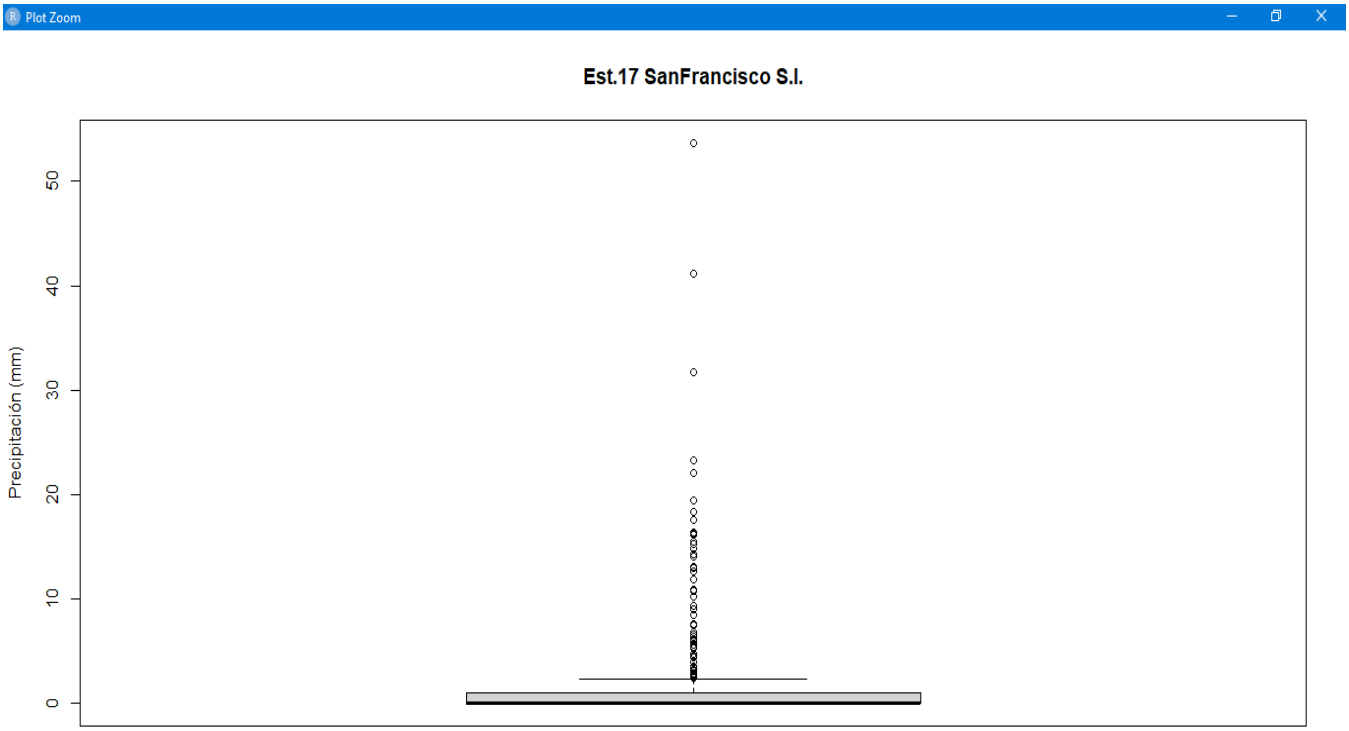


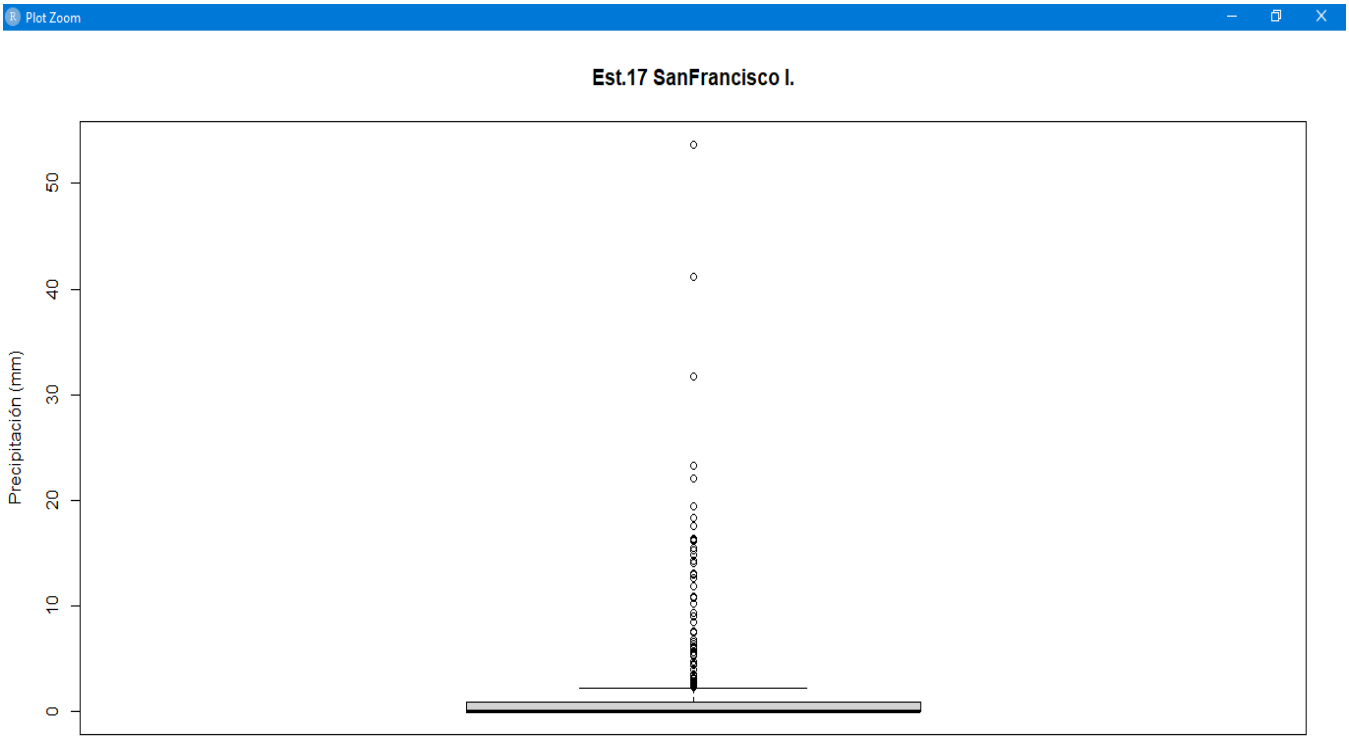
N. 16 Estación NGeneración



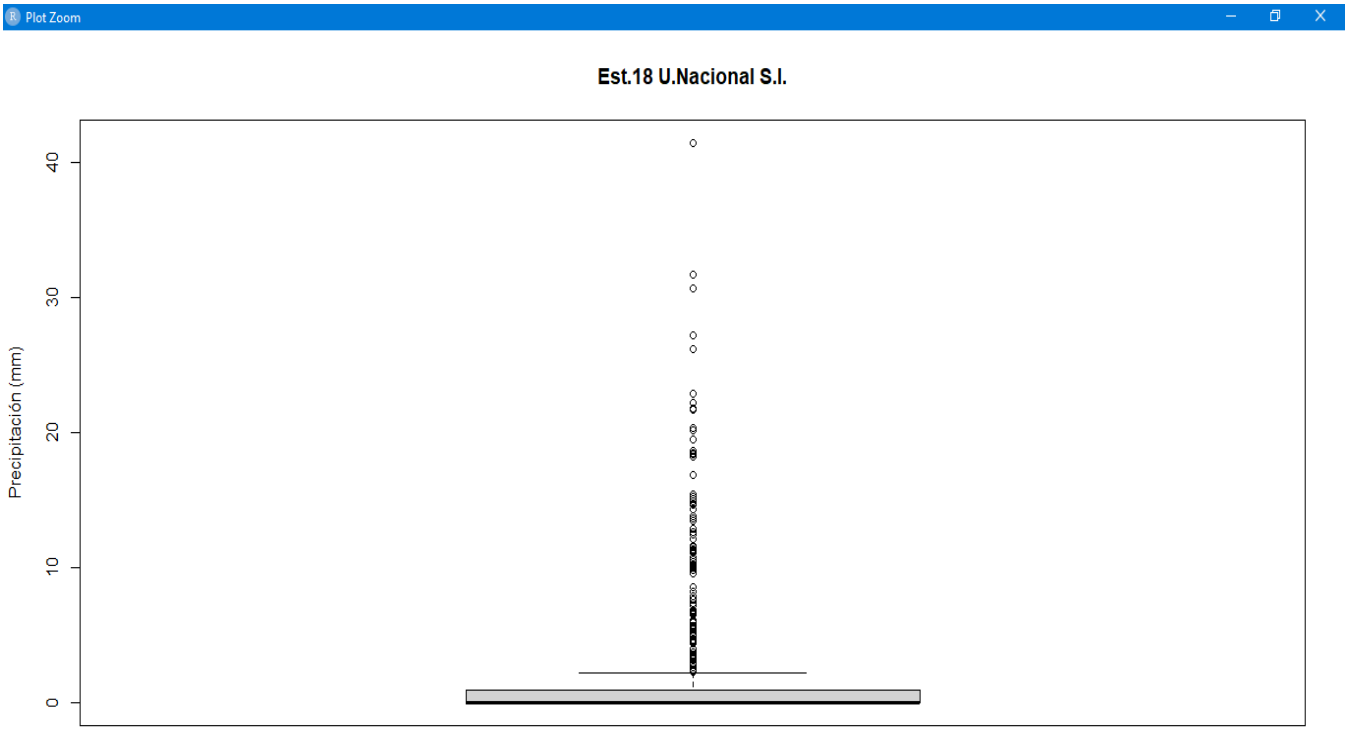


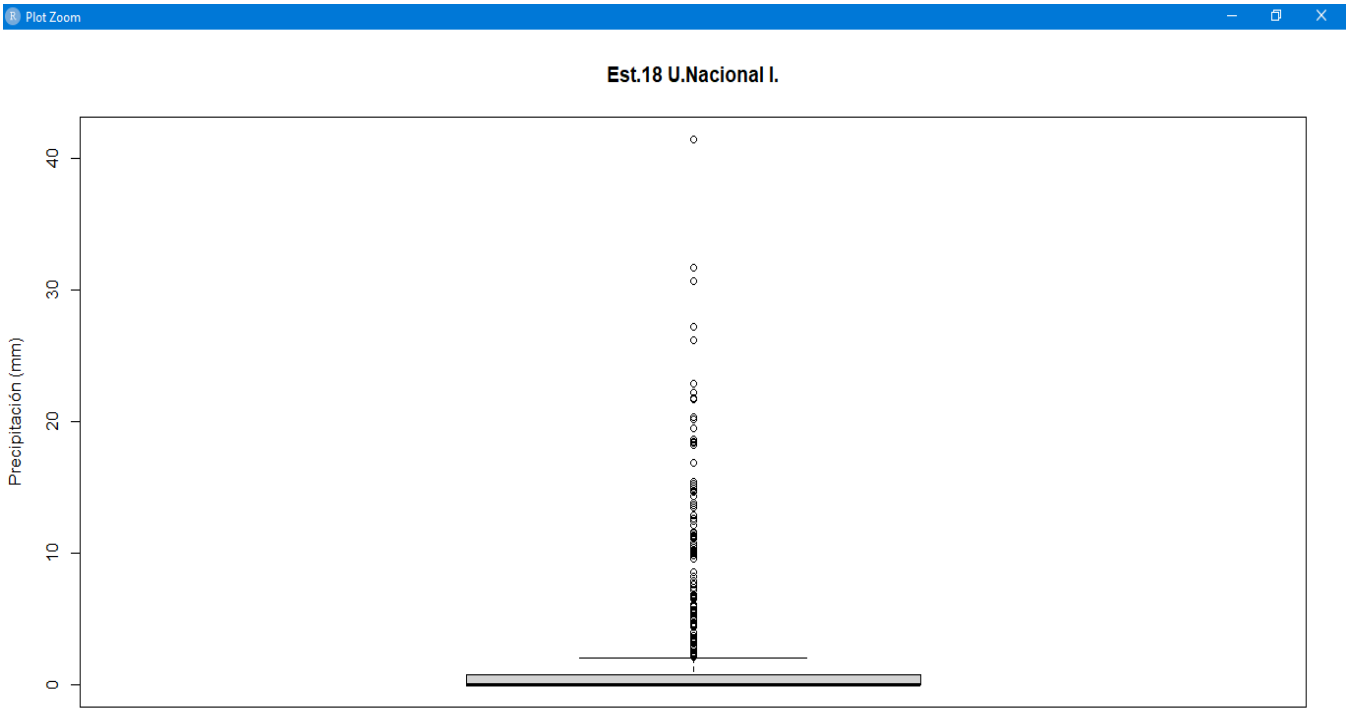
N. 17 Estación S. Francisco





N. 18 Estación U. Nacional





N. 19 Estación VillaTeresa