

# **Reporte de Análisis de Datos de Aberraciones Cromosómicas**

Sergio Daniel Abello Hernández

Universidad Nacional de Colombia Sede Medellín

Fundamentos de Programación para Ciencias Biológicas

2024-2S

# Introducción

## Propósito general del código:

El código está diseñado para realizar análisis exploratorio y visualización de datos sobre anomalías cromosómicas, organizando, limpiando y representando gráficamente la información en múltiples formas. Combina procesamiento de datos, detección de valores atípicos y generación de gráficos para identificar patrones, tendencias y distribuciones relacionadas con anomalías genéticas, cromosómicas y otros atributos asociados.

## Propósitos específicos:

**Procesamiento de datos genéticos:** Se enfoca en limpiar y organizar datos cromosómicos y morfológicos para facilitar su análisis. Y filtra información específica (por ej: elimina cromosomas sexuales X e Y, organiza anomalías y códigos morfológicos).

**Análisis Estadístico y Exploratorio:** Identifica valores atípicos (outliers) en las cuentas de casos utilizando el IQR (Rango Intercuartílico), aplica transformaciones logarítmicas para normalizar datos y facilitar su interpretación visual, y permite detectar qué cromosomas o bandas tienen mayores anomalías o concentraciones de casos.

**Visualización de Datos:** Genera gráficos para comprender la distribución de anomalías y otros parámetros genéticos, como lo son: Gráficos de barras, Mapas de calor, Scatterplots y FacetGrid, facilitando el análisis visual de los datos cromosómicos.

**Creación de un Diccionario de Morfologías:** Extrae y organiza información de códigos morfológicos y nombres de anomalías, creando un diccionario ordenado que puede servir como referencia para análisis posteriores.

## Fuente de información:

Toda la información fue extraída de: Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer. La información de la base de datos Mitelman sobre aberraciones cromosómicas y fusiones de genes en el cáncer relaciona los cambios citogenéticos y sus consecuencias genómicas, en particular las fusiones de genes, con las características de los tumores, ya sea en casos individuales o en asociaciones. Todos los datos han sido seleccionados manualmente de la literatura por Felix Mitelman en colaboración con Bertil Johansson y Fredrik Mertens. La base de datos Mitelman cuenta con el apoyo del Instituto Nacional del Cáncer, la Sociedad Sueca contra el Cáncer y la Fundación Sueca contra el Cáncer Infantil. La base de datos se actualiza trimestralmente en enero, abril, julio y octubre. Específicamente, el DataFrame proviene de la sección de Aberraciones Cromosómicas Recurrentes y en ella, solo se bajaron las aberraciones estructurales para ser analizadas, incluyendo todos los genes, topografías y morfologías y disponibles en la base de datos.

# **Contenido**

1. Página de presentación
2. Introducción
3. Contenido y anexos
4. Índice de gráficas
5. Gráficas finales

# **Índice de Anexos**

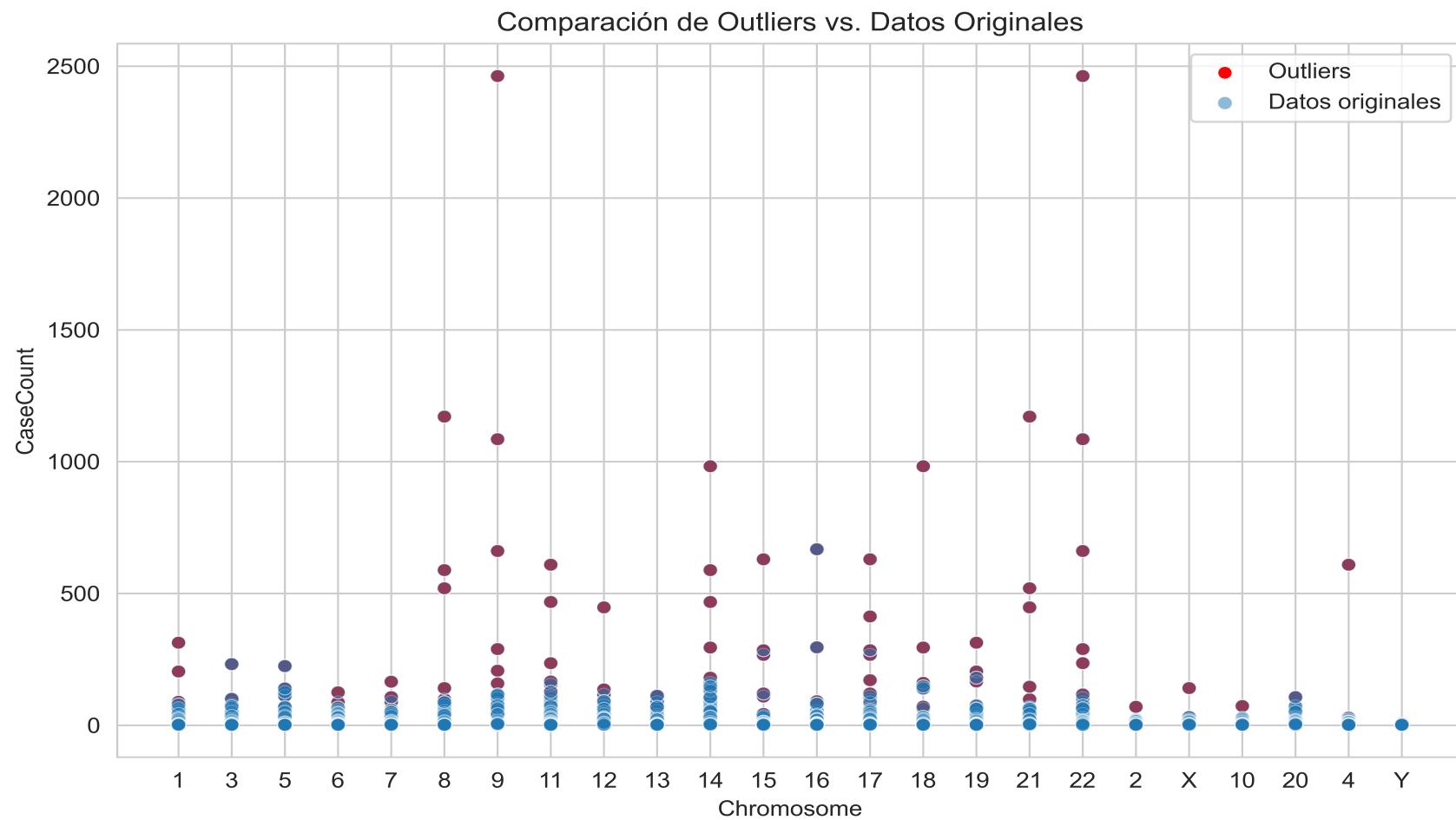
1. DataFrame Original
2. DataFrame Organizado
3. DataFrame Outliers
4. Diccionario de morfologías

# Índice de Gráficas

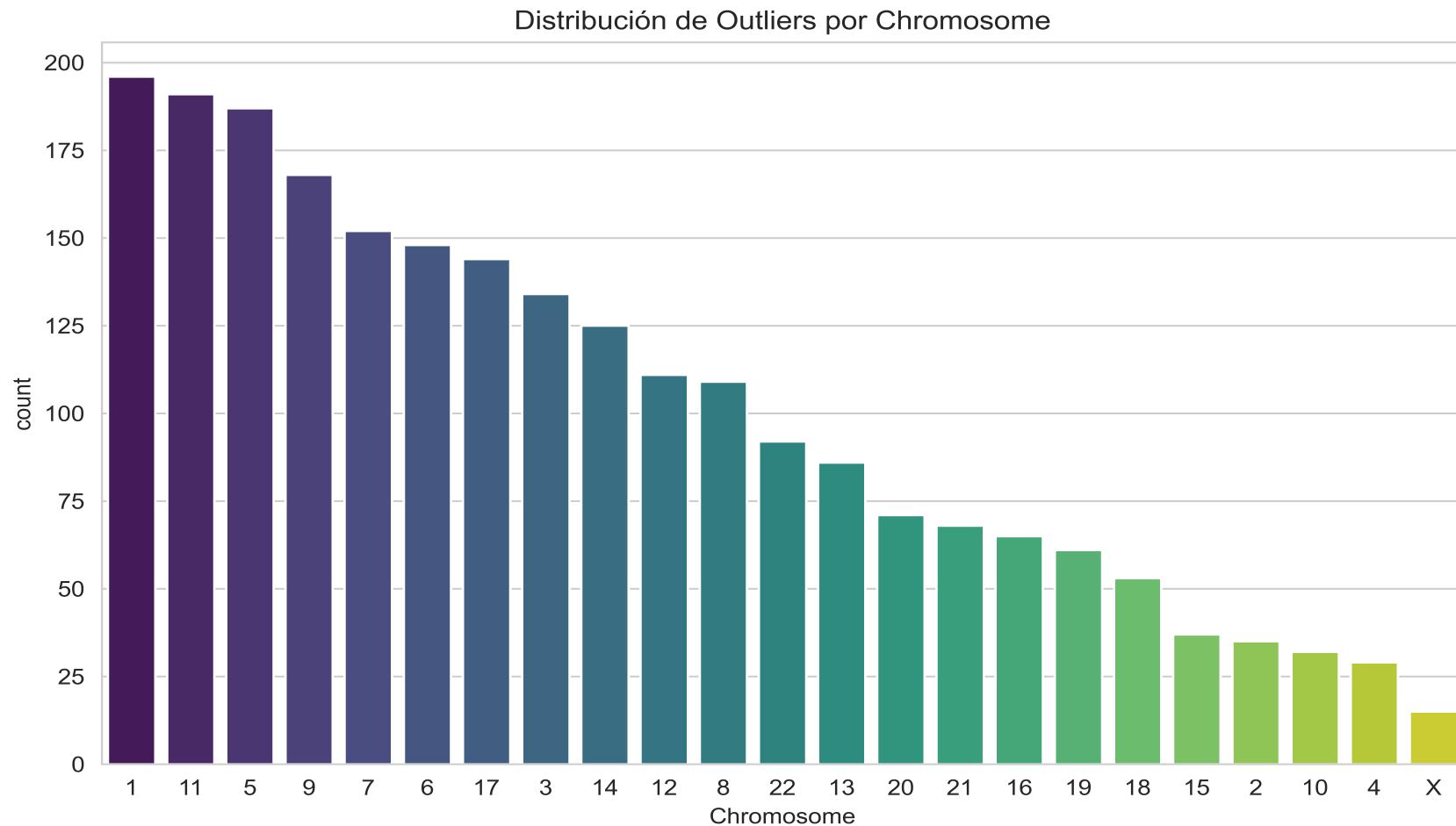
1. Scatterplot: Gráfico de dispersión de outliers
2. Countplot: Gráfico de conteo de outliers por cromosoma
3. Barplot: Conteo de anormalidades por cromosoma y brazo
4. Barplot: Frecuencia de anormalidades
5. Heatmap: Tipo de anormalidad por cromosoma
6. Scatterplot: Distribución de casos por banda cromosómica
7. StripPLOT: Casos por tipo de anormalidad
8. Heatmap: Frecuencia de anormalidades por banda y tipo



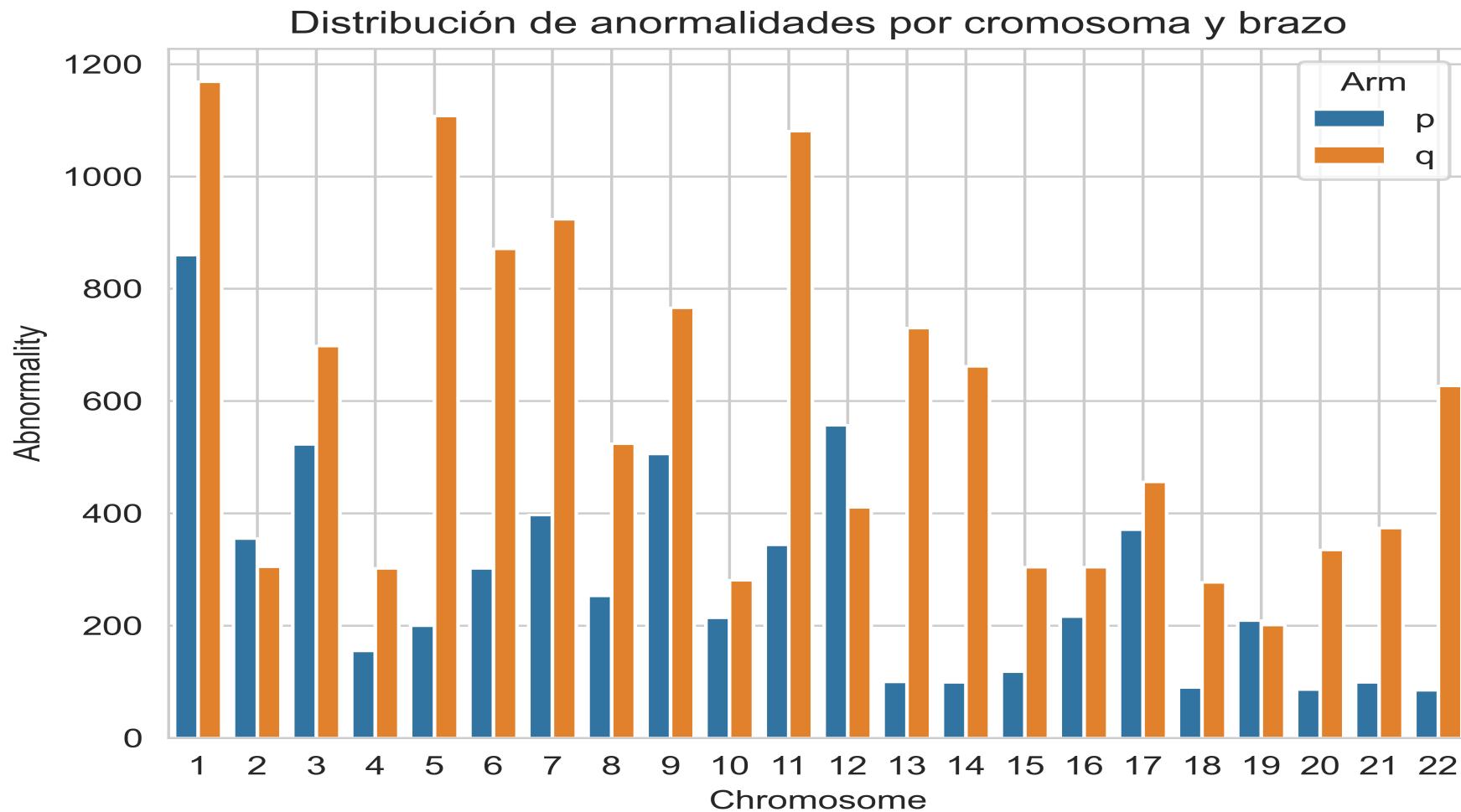
**Gráfica 1**



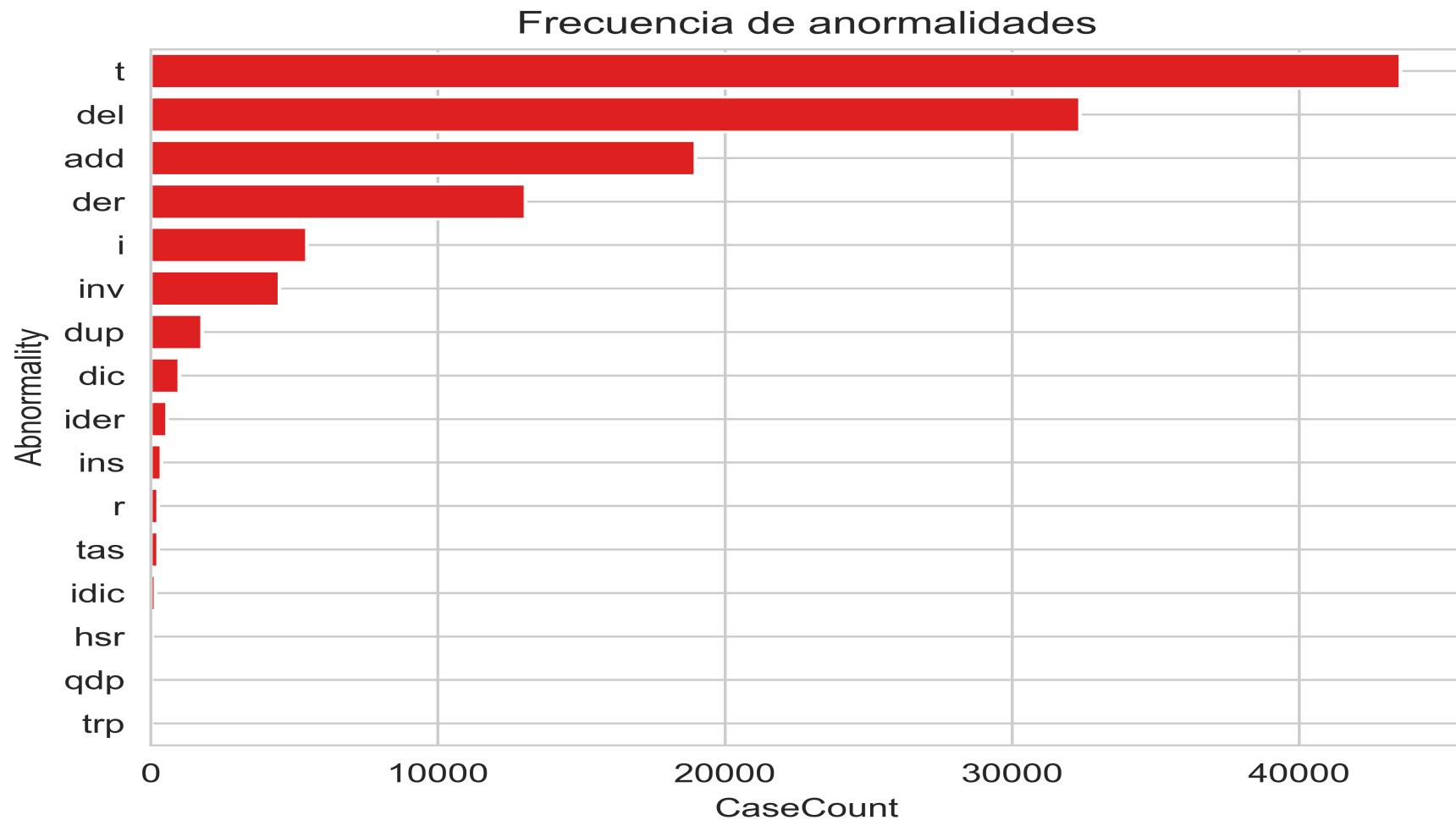
**Gráfica 2**



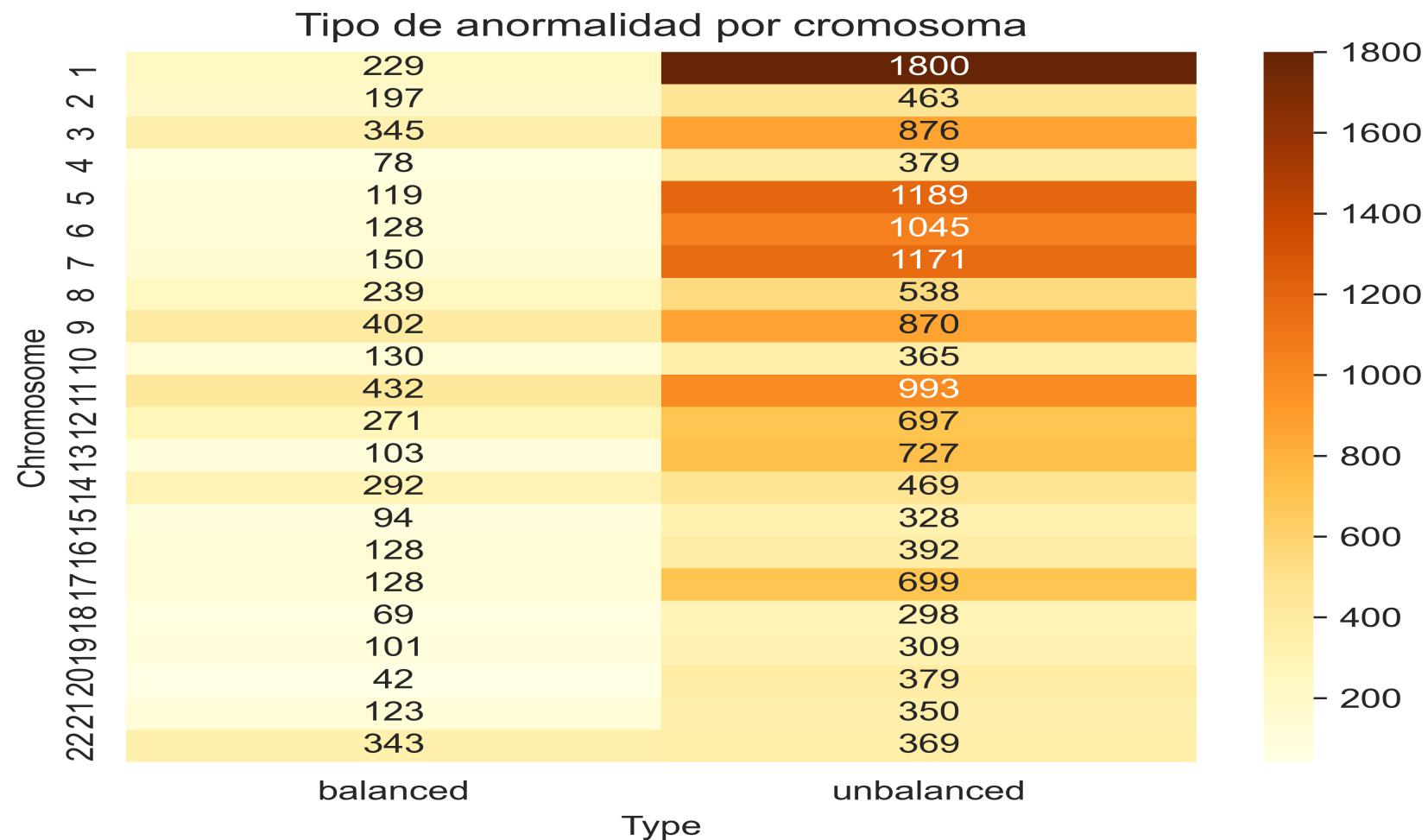
**Gráfica 3**



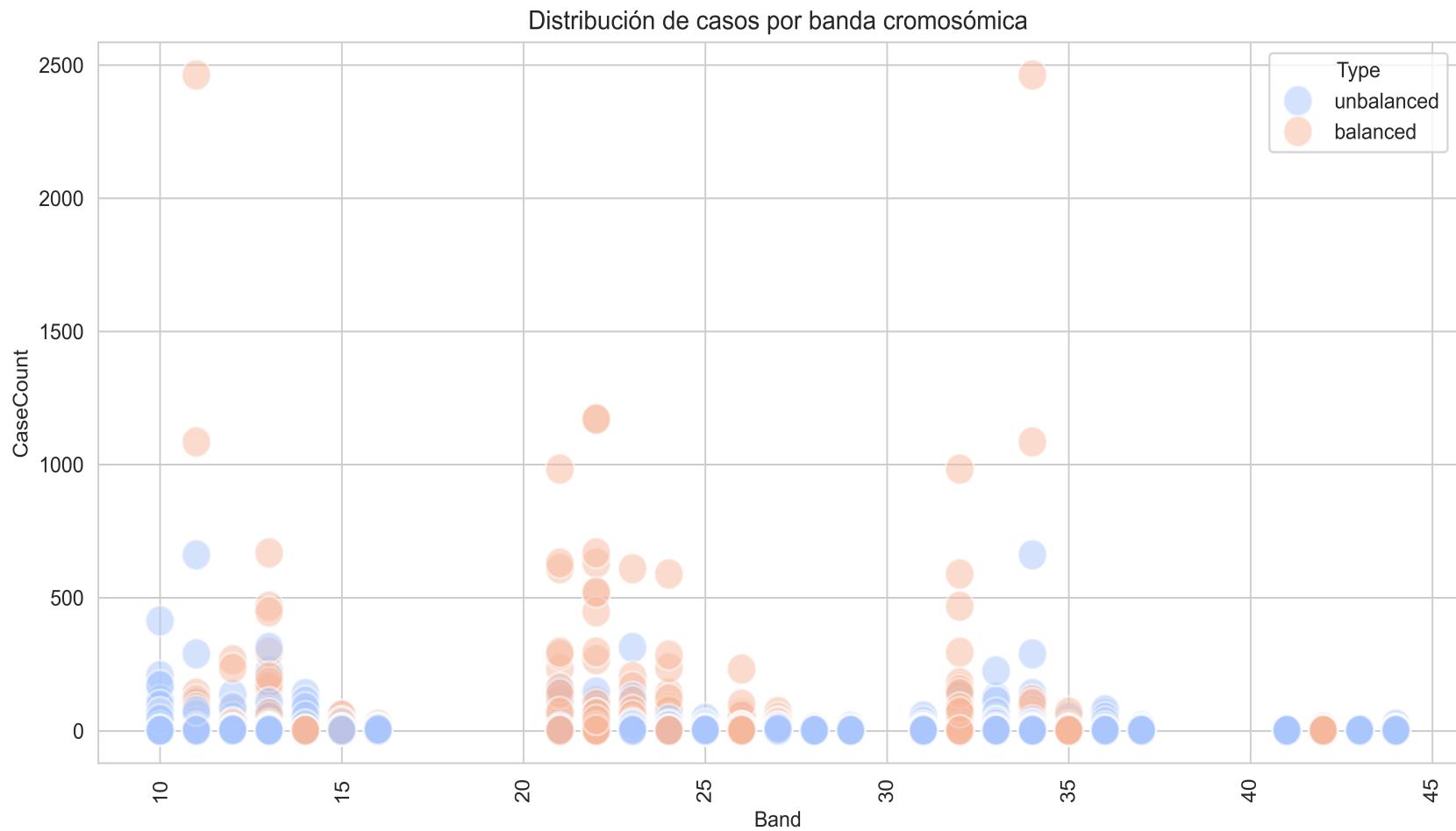
**Gráfica 4**



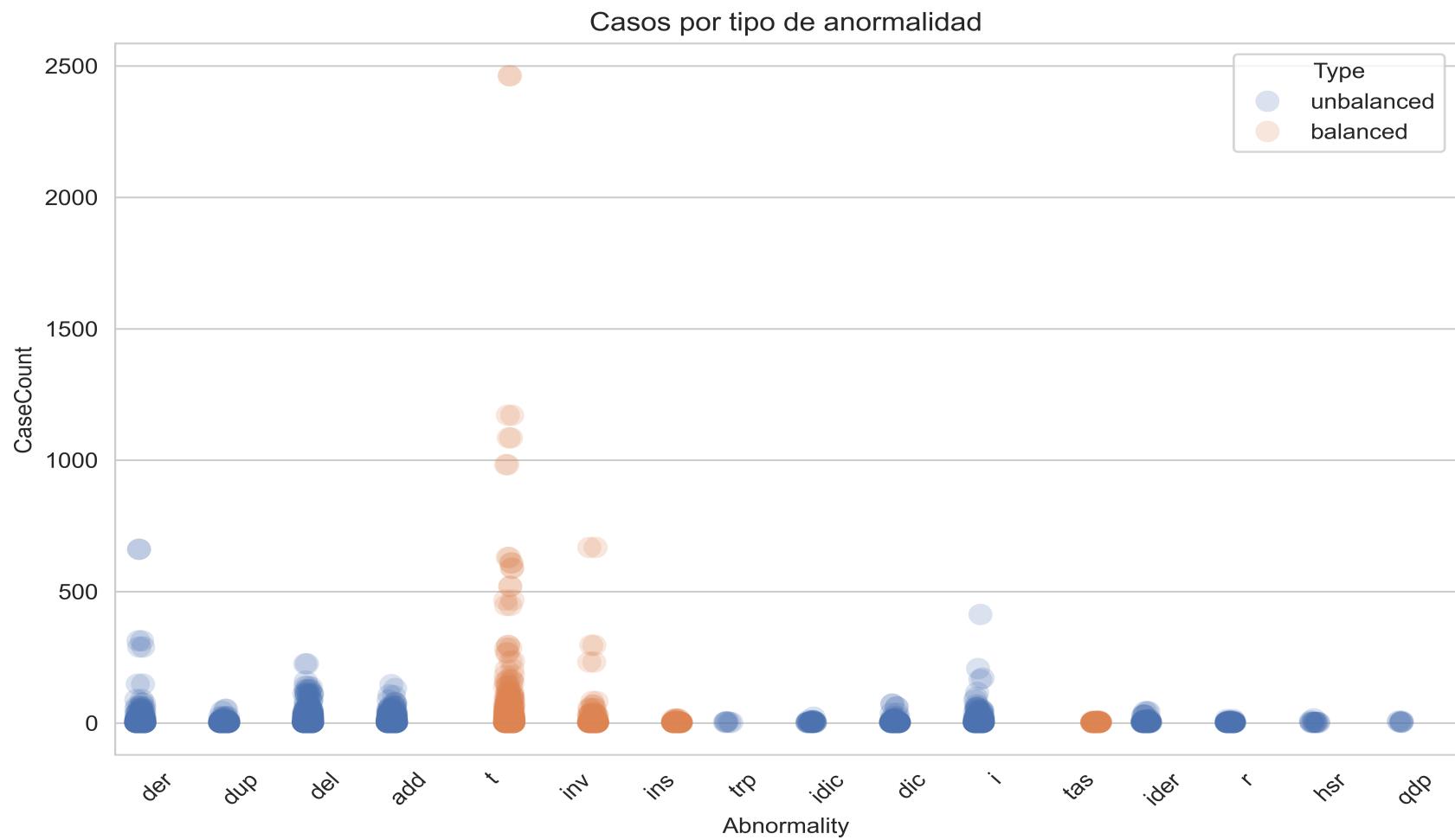
**Gráfica 5**



**Gráfica 6**



**Gráfica 7**



## Gráfica 8

