

Trabajo Práctico nº1

Probabilidad y Estadística

LCC

Augusto Rabbia
Manuel Spreutels
Agustín Torres



Julio 2023

1 Introducción

En el año 2011, la Ciudad de Buenos Aires realizó un censo forestal urbano público, con el objetivo de contabilizar y determinar el estado actual del arbolado urbano público.

Se estudiará una muestra extraída de este censo.

2 Objetivo

El objetivo es resumir la información y dar claridad sobre los datos, resaltando las características más notables que estos presentan.

3 Desarrollo

3.1 Análisis univariado

El conjunto de datos consta de 350 árboles en la Ciudad de Buenos Aires, dando información sobre sus siguientes atributos:

Nombre	Descripción
ID	Identificación del árbol.
altura	Altura de cada árbol, medida en metros (m). Observación: si un árbol mide 12,7 m se tomará como dato "12", truncando los valores a la unidad.
diámetro	Diámetro de cada árbol, medido en centímetros (cm).
inclinación	Ángulo que forma el tronco del árbol respecto a una perpendicular al suelo, medido en grados (°). Indica el grado de inclinación del árbol.
especie	Especie a la que pertenece el árbol, dentro de las siguientes categorías: Eucalipto, Jacarandá, Palo borracho, Casuarina, Fresno, Ceibo, Ficus, Álamo, Acacia.
origen	Procedencia de la especie: Exótico, Nativo/Autóctono, No Determinado.
brotos	Número de brotes jóvenes crecidos durante el último año.

Un rasgo que puede ser de interés para describir el arbolado urbano de la ciudad es la distribución de este según su especie:

Table 1: Tabla de frecuencias de la variable especie

Especie	Frec. absoluta	Frec. relativa	Frec. absoluta	Frec. relativa acumulada
Acacia	23	0.07	23	0.07
Álamo	73	0.21	96	0.28
Casuarina	50	0.14	146	0.42
Ceibo	14	0.04	160	0.46
Eucalipto	68	0.19	228	0.65
Ficus	13	0.04	241	0.69
Fresno	28	0.08	269	0.77
Jacarandá	39	0.11	308	0.88
Palo borracho	42	0.12	350	1
Total	350	1		

La tabla, como se puede observar, es heterogénea respecto a la frecuencia que presenta cada especie. En efecto, el árbol más común, es decir, la moda, es el Álamo, que llega a representar más de $1/5$ de la muestra, con un 21% del total de los árboles, mientras que el Ceibo y el Ficus no representan $1/20$ del total individualmente, siendo cada uno un 4% del total.

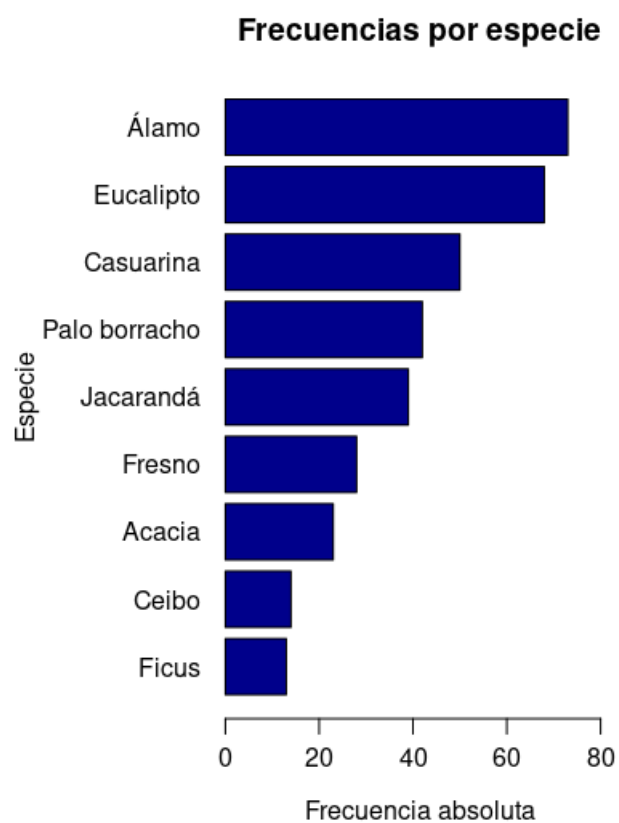


Figure 1: Gráfico de Barras de la Variable Especie

La distribución del origen de los árboles también resulta relevante a la hora de analizar el arbolado.

Table 2: Tabla de frecuencias de la variable origen

Origen	Frec. absoluta	Frec. relativa	Frec. absoluta	Frec. relativa acumulada
Exótico	251.00	0.72	251.00	0.72
Nativo/Autóctono	99.00	0.28	350.00	1.00
Total	350	1		

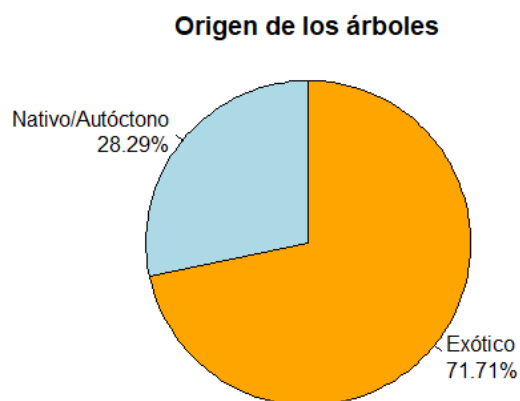


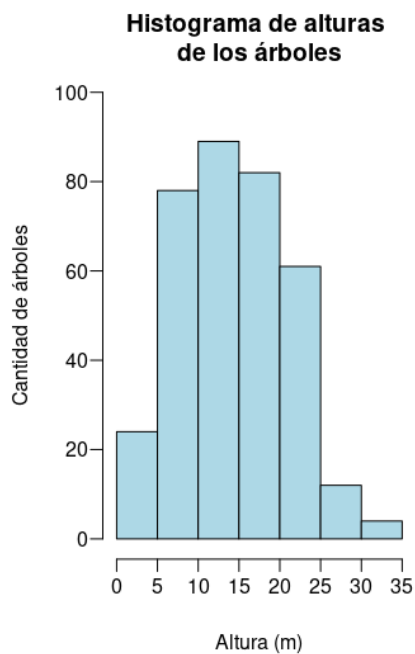
Figure 2: Proporciones de los orígenes de los árboles.

Se observa que la cantidad de árboles de origen exótico en la muestra analizada más que duplica a la de aquellos de origen nativo o autóctono, por lo que la moda de origen es origen exótico. Esto parece indicar que el origen de un árbol no parece ser un factor a la hora de decidir qué árbol plantar, por lo contrario, elegir un arbol de nativo parece ser un rasgo de relevancia mínima, optando en cambio por cualidades que serán vistas a continuación.

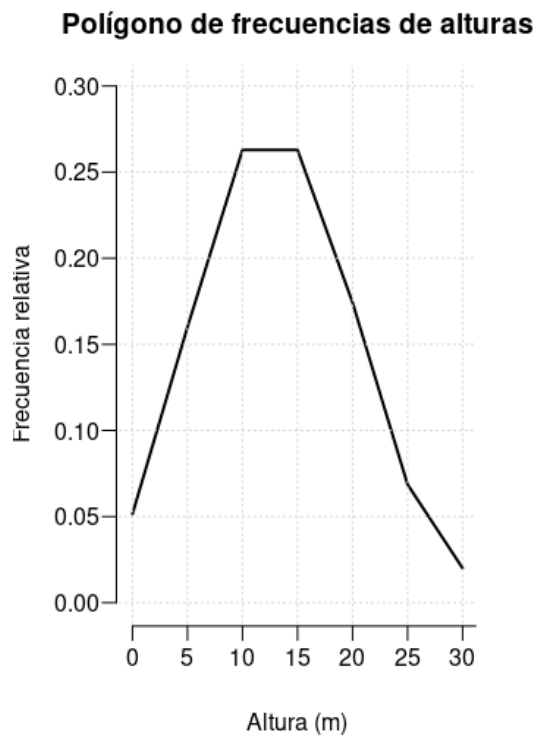
Otra variable de interés para el análisis descriptivo es la altura de los árboles.

Table 3: Tabla de frecuencias de la variable altura

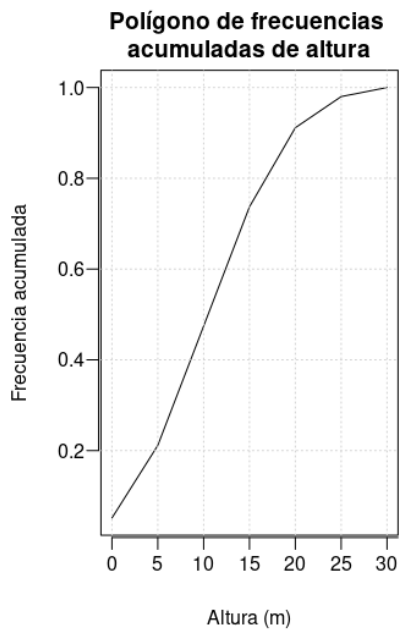
Altura (m)	Frec. absoluta	Frec. relativa	Frec. absoluta	Frec. relativa acumulada
[0,5]	24.00	0.07	24.00	0.07
(5,10]	78.00	0.22	102.00	0.29
(10,15]	89.00	0.25	191.00	0.55
(15,20]	82.00	0.23	273.00	0.78
(20,25]	61.00	0.17	334.00	0.95
(25,30]	12.00	0.03	346.00	0.99
(30,35]	4.00	0.01	350.00	1.00
Total	350	1		



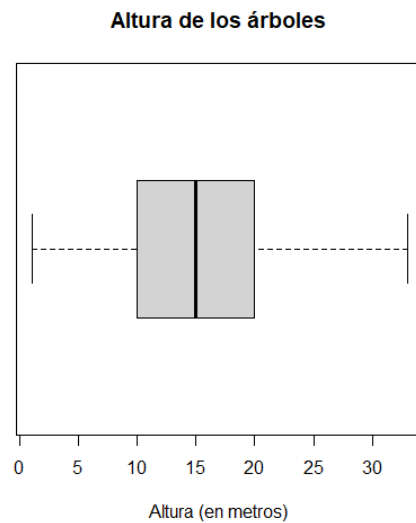
(a) Histograma de la variable altura.



(c) Frecuencias relativas de altura.



(b) Frecuencias relativas acumuladas de altura.



(d) Boxplot de la variable altura.

Figure 3: Análisis univariado de la altura

El promedio de la altura de los árboles de la muestra es de 14.85m, lo cual coincide con la mediana, que tiene un valor de 15m. La desviación estándar es de 6,61m. Además se puede observar que no hay presencia de outliers en el conjunto de datos. Otros datos de interés es que hay 2 árboles de mayor altura, con 33 metros, ambos de la especie Eucalipto.

El primer cuartil tiene un valor de 10m, y el tercero un valor de 20m. El mínimo para esta variable es de 1m. Es probable que esto se deba a que esta variable está discretizada en el conjunto de datos o a que los árboles menores a un metro no hayan sido considerados para obtener la muestra.

Table 4: Tabla de frecuencias de la variable inclinación

Inclinación	Frec. absoluta	Frec. relativa	Frec. absoluta	Frec. relativa acumulada
[0,5]	269.00	0.77	269.00	0.77
(5,10]	48.00	0.14	317.00	0.91
(10,15]	21.00	0.06	338.00	0.97
(15,20]	8.00	0.02	346.00	0.99
(20,25]	2.00	0.01	348.00	0.99
(25,30]	1.00	0.00	349.00	1.00
(30,35]	0.00	0.00	349.00	1.00
(35,40]	0.00	0.00	349.00	1.00
(40,45]	1.00	0.00	350.00	1.00
Total	350	1		

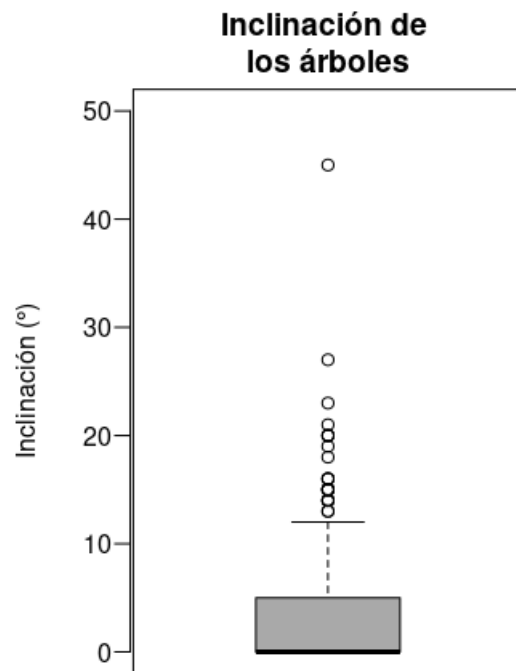


Figure 4: Boxplot de la variable inclinación

En la [Tabla 4](#) se aprecia que más de tres de cada cuatro árboles presenta una inclinación leve ($\leq 5^\circ$) o nula, lo cual podría ser indicativo de que se hace un buen mantenimiento en términos de extracción de árboles peligrosos, ya que un árbol inclinado puede ser más propenso a caídas. Sin embargo, se puede observar en la Figura 4 la notoria presencia de outliers para esta variable, con un rango intercuartílico de 5° . La desviación estándar para la variable es de $5,5^\circ$, y el promedio es de $2,92^\circ$. La inclinación mínima coincide con la mediana, y es de 0° , indicando un árbol perpendicular al suelo, y la máxima es de 45° , habiendo un único ejemplar con esta inclinación en la muestra, y siendo además un outlier remoto.

El diámetro es una variable que puede brindar abundante información sobre los árboles en estudio:

Table 5: Tabla de frecuencias de la variable diámetro

Diámetro	Frec. absoluta	Frec. relativa	Frec. absoluta	Frec. relativa acumulada
[0,18]	60	0.17	60	0.17
(18,36]	144	0.41	204	0.58
(36,54]	62	0.18	266	0.76
(54,72]	45	0.13	311	0.89
(72,90]	21	0.06	332	0.95
(90,108]	9	0.03	341	0.97
(108,126]	4	0.01	345	0.99
(126,144]	1	0	346	0.99
(144,162]	2	0.01	348	0.99
(162,180]	2	0.01	350	1
Total	350	1		

Se puede contemplar en la [Tabla 5](#) que casi 9 de cada 10 árboles tienen un diámetro menor o igual a 70 centímetros.

La desviación estándar es de 27,78cm por lo que, dado el rango que recorre la variable, da lugar a gran cantidad de outliers, lo cual es apreciable en la [Figura 5](#). El promedio es de 40,82cm, con un rango intercuartilíco de 22cm. La mediana es de 33cm y el tercer cuartil es de 53cm. Entre los outliers, se destacan un árbol con 180cm de diámetro, siendo este el máximo. Por otro lado, el diámetro mínimo registrado es de 2cm.

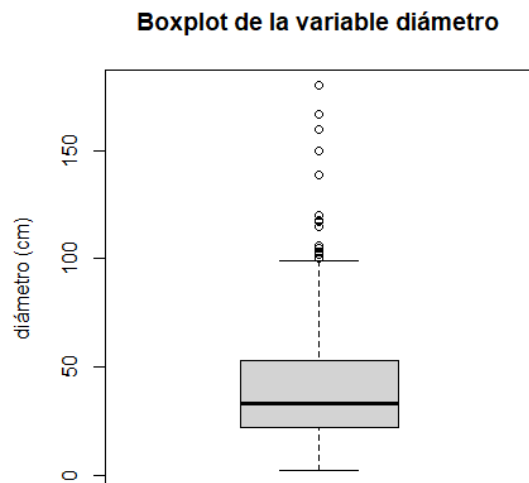


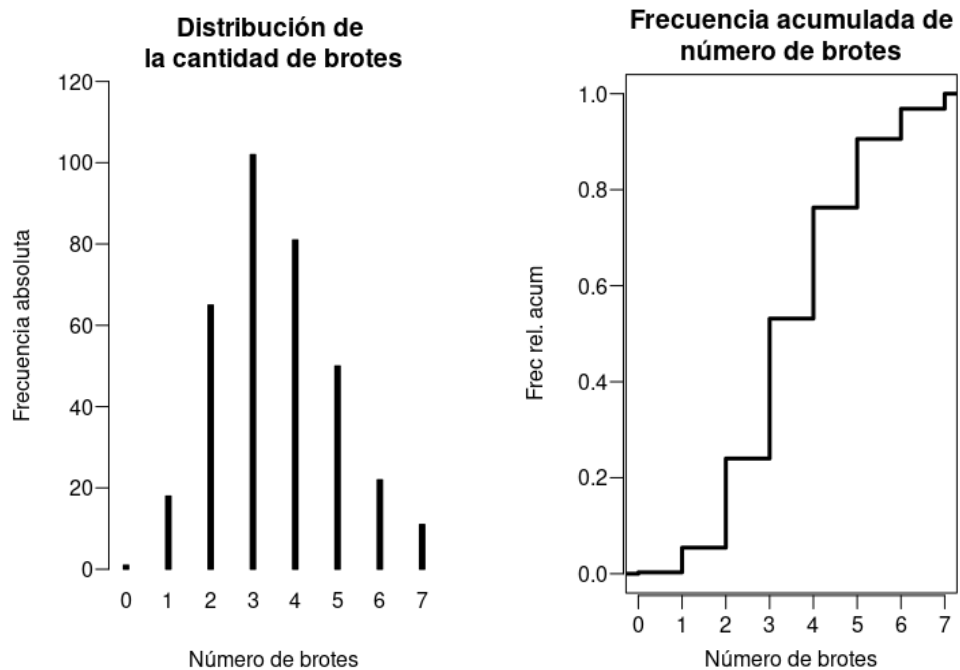
Figure 5: Boxplot de la variable diámetro.

Finalmente, la cantidad de brotes que posean los árboles puede ser un buen indicador de la vejez del arbolado, permitiendo hacer predicciones sobre su evolución en el tiempo.

Table 6: Tabla de frecuencias de la variable brotes

Brotes	Frec. absoluta	Frec. relativa	Frec. absoluta	Frec. relativa acumulada
0	1.00	0.00	1.00	0.00
1	18.00	0.05	19.00	0.05
2	65.00	0.19	84.00	0.24
3	102.00	0.29	186.00	0.53
4	81.00	0.23	267.00	0.76
5	50.00	0.15	317.00	0.91
6	22.00	0.06	339.00	0.97
7	11.00	0.03	350.00	1.00
Total	350	1		

La variable tiene un rango de 0 a 7, con un 75% de la muestra con 4 brotes o menos, y un rango intercuartílico de 1, puesto que su primer cuartil es de 3 brotes. La desviación estándar que presenta es de 1.41 brotes. La mediana se ubica en 3 brotes, mientras que el promedio, ligeramente mayor, es de 3.53 brotes.



(a) Distribución de brotes por árbol.

(b) Proporciones de números de brotes

Figure 6: Análisis del número y distribución de brotes

3.2 Análisis bivariado

A continuación, se explorarán algunas relaciones existentes en los datos.

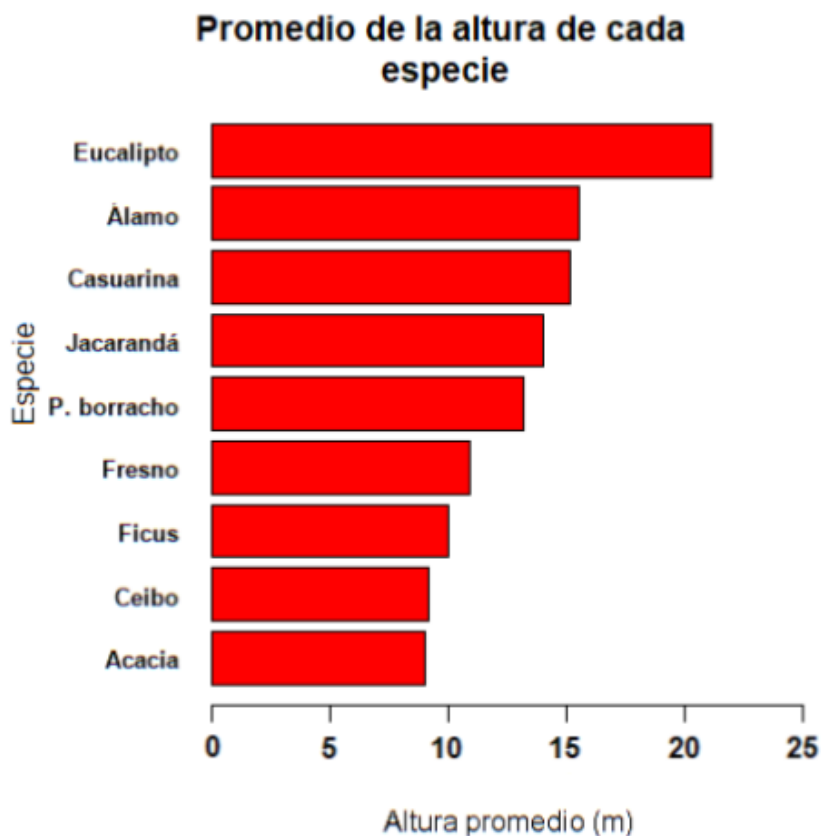
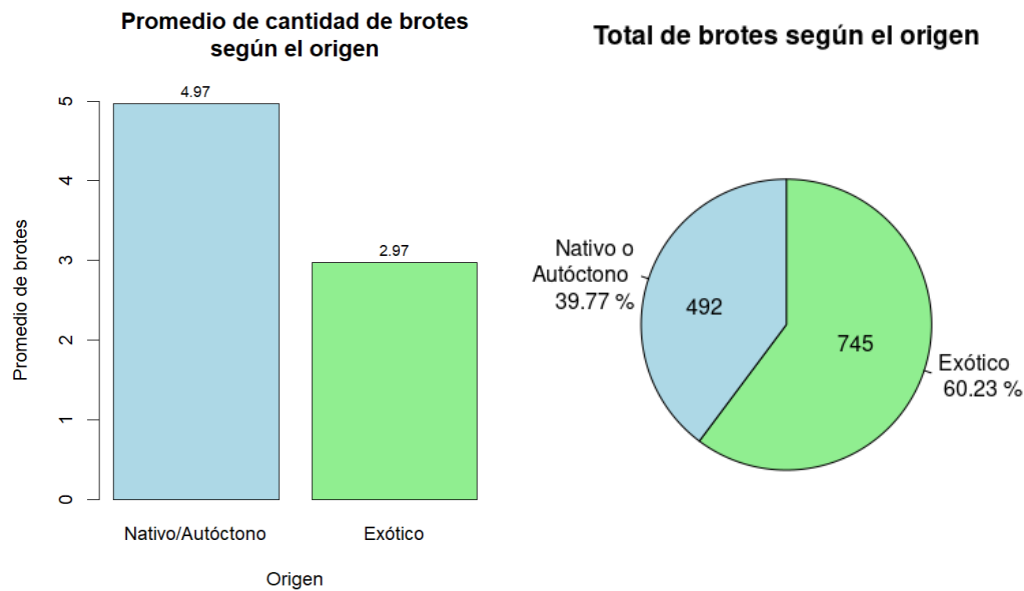


Figure 7: Promedio de la altura de cada especie.

Al analizar la altura de los árboles según su especie, resalta el hecho de que 6 de las 9 especies poseen una altura promedio menor al promedio global, de 14.85m, así como la mediana, de 15m, como vimos en la sección de [análisis univariado](#). Esto se debe a que las más altas en promedio son las más comunes en la muestra, como se observa en la [Tabla 1](#). Este último fenómeno puede explicarse cuando consideramos la demanda de los árboles, pues en general, suele apreciarse más por los compradores el tener un árbol de mayor altura, por razones entre las que se encuentran cualidades como la cantidad de sombra que producen o por preferencias estéticas.



(a) Cantidad de árboles según origen por brotes.

(b) Total de los brotes según el origen.

Figure 8: Análisis bivariado de los brotes

Al estudiar la prosperidad de los árboles en base a su origen, se observa que existe una tendencia entre los árboles nativos a tener más brotes que los exóticos. En la [Figura 8 \(a\)](#), donde se observa que, en promedio, un árbol de origen nativo/autóctono tiene 2 brotes más que uno exótico. Estos números significan que los arboles de origen nativo tienen 67.34% más brotes que sus pares de origen exótico.

Estos resultados resaltan de forma aún más llamativa al examinar la proporción de árboles nativos comparada con la de exóticos en la [Figura 2](#) y la de brotes de árboles exóticos y nativos que se observan en la [Figura 8 \(b\)](#). Se ve que la proporción entre estos no sólo no se mantiene, sino que existe una diferencia de más de 10 puntos porcentuales.

Esto indica que el crecimiento de los árboles nativos es más rápido que los exóticos. Esto puede deberse al hecho que estos están mejores adaptados al suelo y las condiciones climáticas locales.

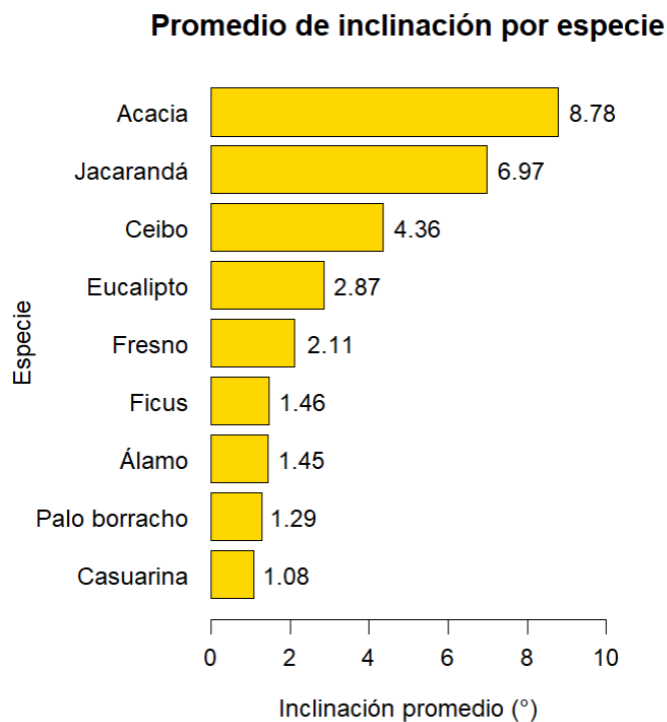


Figure 9: Promedio de la inclinación de cada especie en grados.

Se ve que hay una gran diferencia en la inclinación promedio de cada especie. La inclinación promedio mínima la tienen los árboles de especie Casuarina, con 1.08° , mientras que la máxima es más de 8 veces mayor, y la tienen los árboles de Acacia, con 8.78° .

Sin embargo, puede resultar importante considerar que únicamente hay 23 árboles de Acacia en la muestra, como se puede ver en la [Tabla 1](#). Esto puede ser un factor que explique el valor tan alto en esta variable. Ahora bien, la baja presencia de árboles de especie Acacia puede ser explicada, justamente, por la alta inclinación este árbol tiende a tener, pues esto podría hacerlo una opción menos atractiva para posibles compradores, o podría llevar a que los árboles con mayor inclinación sean removidos.

Durante el análisis, sin embargo, destacó el hecho que el árbol de mayor inclinación en la muestra, un outlier remoto con 45° de inclinación, es uno de esta especie, cosa que podría ser un factor a considerar a la hora de analizar este resultado.

4 Conclusiones

Se puede resumir las características del conjunto de datos analizado en los siguientes puntos:

- Se observa una heterogeneidad en la frecuencia de las especies de árboles en la muestra. El árbol más común es el "Álamo", representando más de 1/5 de la muestra, seguido por el "Eucalipto". En contraste, especies como el "Ceibo" y el "Ficus" tienen una representación mínima.
- La cantidad de árboles de origen exótico duplica la de árboles de origen nativo. Parece indicar que el origen del árbol no es un factor determinante al decidir qué árbol plantar en la ciudad.
- La altura promedio de los árboles de la muestra es de 14.85m, y la mediana es de 15m, sin presencia de outliers en los datos.
- El 98% de los árboles tuvieron al menos un brote joven durante el año anterior a la toma de la muestra, lo que sugiere que incluso las especies exóticas gozan de buenas condiciones en el suelo sobre el que están. También podría ser indicativo de la juventud de los árboles que conforman el conjunto de datos.
- Hay dos árboles de mayor altura con una altura de 33 metros cada uno, ambos de la especie Eucalipto, que es también la especie más alta en promedio.
- La cantidad de árboles de origen exótico en la muestra duplica a la de aquellos de origen nativo o autóctono.
- Aproximadamente el 90% de los árboles tienen un diámetro menor o igual a 70 centímetros. Se observa la presencia de outliers en esta variable, con un rango amplio de diámetros.
- Las especies más comunes son también las más altas en promedio, sugiriendo una preferencia de los consumidores por tener árboles de mayor altura, sea esto por razones estéticas, de comodidad, u otras cualidades como la cantidad de sombra que estos ofrecen.
- La inclinación es una variable donde los outliers pueden representar un peligro, al poder ser un indicio de que un árbol tenga un mal balance o pueda caerse. Pero es una variable donde los outliers son infrecuentes, indicando un buen cuidado entre los árboles de la Ciudad.

- La inclinación promedio parece ser una característica inherente de la especie, y no de la ubicación o las condiciones en las que está plantado un árbol.
- La tendencia de los árboles nativos a tener más brotes que los exóticos sugiere que su fertilidad es mayor, lo que puede deberse a su mejor adaptación al suelo y las condiciones climáticas locales.

En resumen, la altura, la especie, el origen, el diámetro y la cantidad de brotes jóvenes son datos importantes para evaluar la salud y el valor ecológico del arbolado urbano. Al comprender estos datos y utilizarlos en la gestión del arbolado, es posible mejorar la calidad de vida de las personas en las ciudades y promover un entorno más saludable y sostenible.