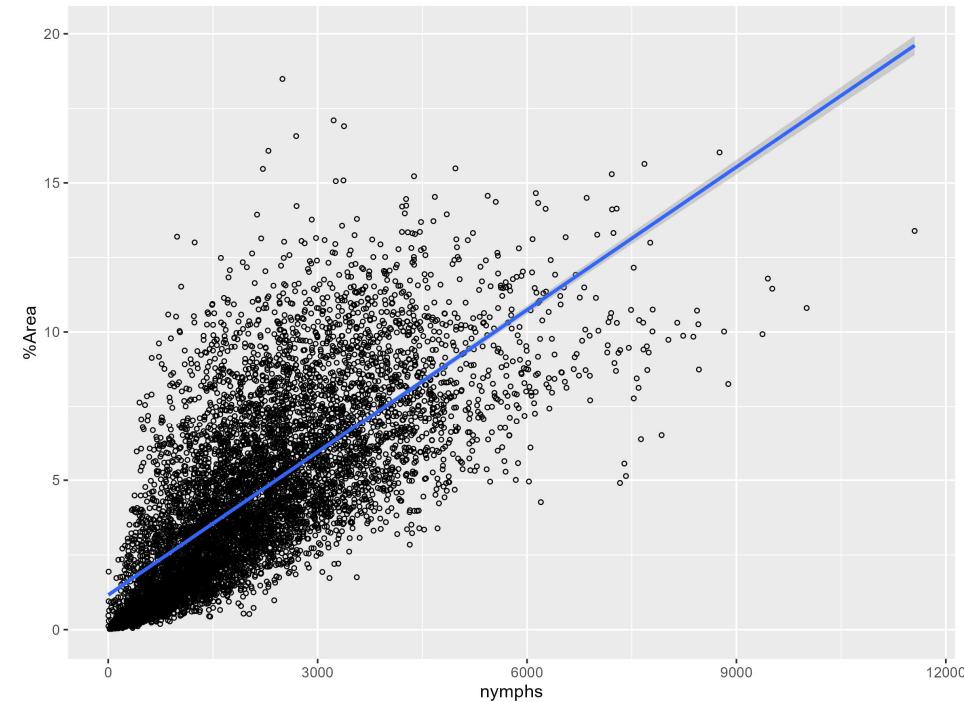


# Representación grafica de datos cuantitativos

Luis Fernando Delgado Muñoz  
Ingeniero Agroindustrial  
Magister en Ciencias Agrarias

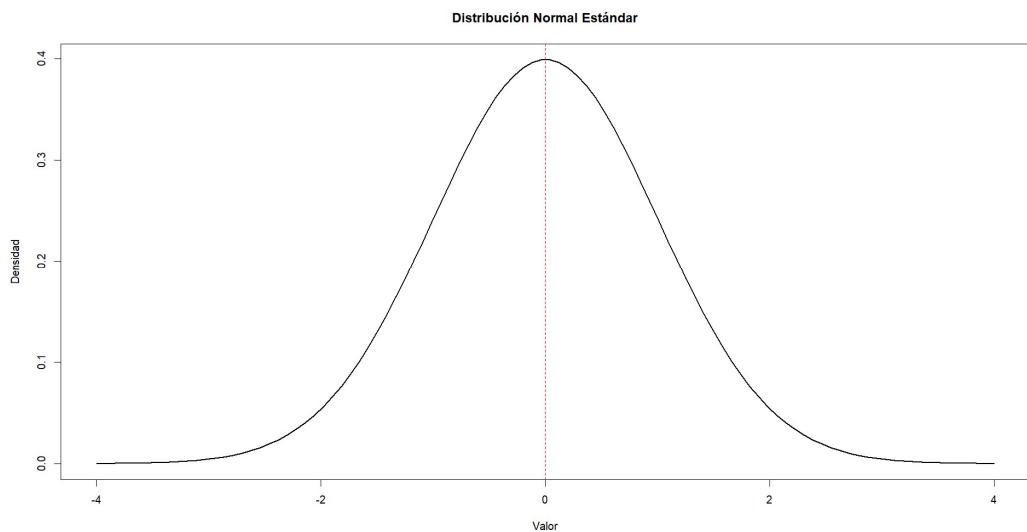


# Medidas de tendencia central

- **Moda:** El valor o valores más frecuentes en un conjunto de datos.
- **Mediana:** El valor medio de un conjunto de datos ordenado. Si hay un número par de datos, es el promedio de los dos valores centrales.
- **Media:** El promedio de un conjunto de datos, calculado sumando todos los valores y dividiéndolos por el número total de valores.
- **Ejemplo:** Calcule los indicadores de tendencia central para el siguiente conjunto de datos:
- Conjunto de datos:

5 6 6 7 7 7 8 8 9 10

5 6 6 7 7 7 8 8 9 20



# Indicadores de dispersión

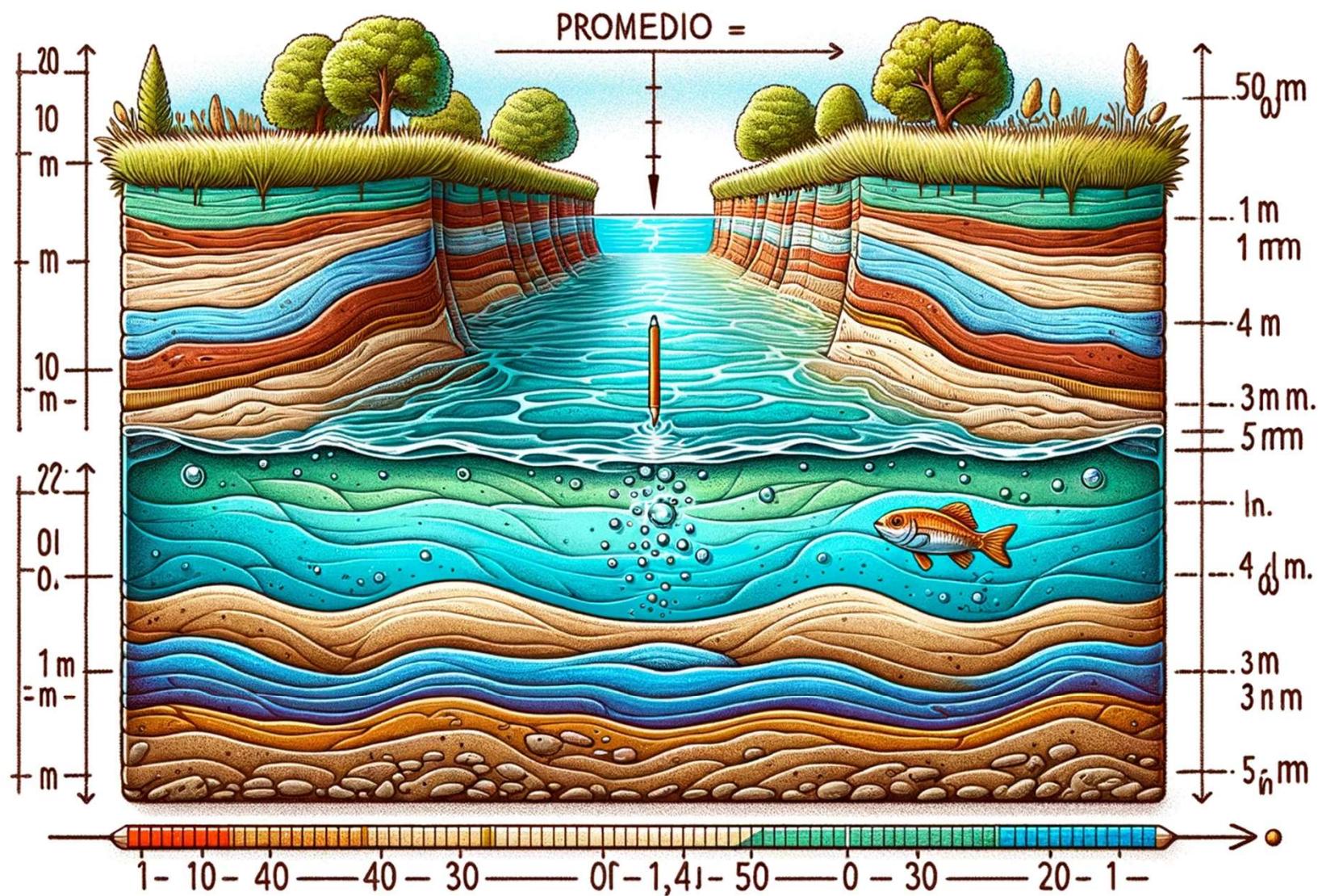
- Los indicadores de tendencia central, como la **media**, **mediana** y **moda**, proporcionan un resumen del punto central de un conjunto de datos. Sin embargo, para obtener una visión completa de la distribución de los datos, deben complementarse con medidas de dispersión.

## Importancia de las Medidas de Dispersión:

- Un índice de dispersión reducido señala que los valores de un conjunto de datos están cercanos entre sí, lo que confiere mayor confiabilidad al indicador de tendencia central escogido.
- Por el contrario, un índice de dispersión elevado revela que hay una amplia variabilidad en los datos, lo que podría disminuir la confiabilidad de un solo valor de tendencia central para representar todo el conjunto de datos.



Profundidad promedio de 1.40 mts.  
¿Cruzaria el rio?



### *Genotypes evaluated*

Genotypes	Average optimal cooking time (Feb-2021) (min)	Standard deviation (min)
BRA158	28	13
IND129	24	5
CM6370-2	33	14
VEN117B	24	16
COL1516	27	13
PER368	28	19
MAL3	25	9
CM5948-1	31	13
CM7436-7	29	9
COL2627	33	12
VEN77	33	13
CUB46	22	6
COL2246	38	18



## Rango

$$R = \max - \min$$

## Desviación Media

$$DM = \frac{\sum_{i=1}^N |x_i - \bar{x}|}{N}$$

donde:

- $N$  es el número de observaciones,
- $x_i$  es cada observación individual,
- $\bar{x}$  es la media aritmética de todas las observaciones.

## Standard deviation (population)

$$S = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

donde:

- $N$  es el número de observaciones en la población.
- $x_i$  es cada observación individual,
- $\mu$  es la media de la población.

## Standard deviation (sample)

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

donde:

- $N$  es el tamaño de la muestra.
- $x_i$  es cada observación individual,
- $\bar{x}$  es la media de la muestra.

## Variance (population)

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

donde:

- $N$  es el número de observaciones en la población.
- $x_i$  es cada observación individual,
- $\mu$  es la media de la población.

## Variance (sample)

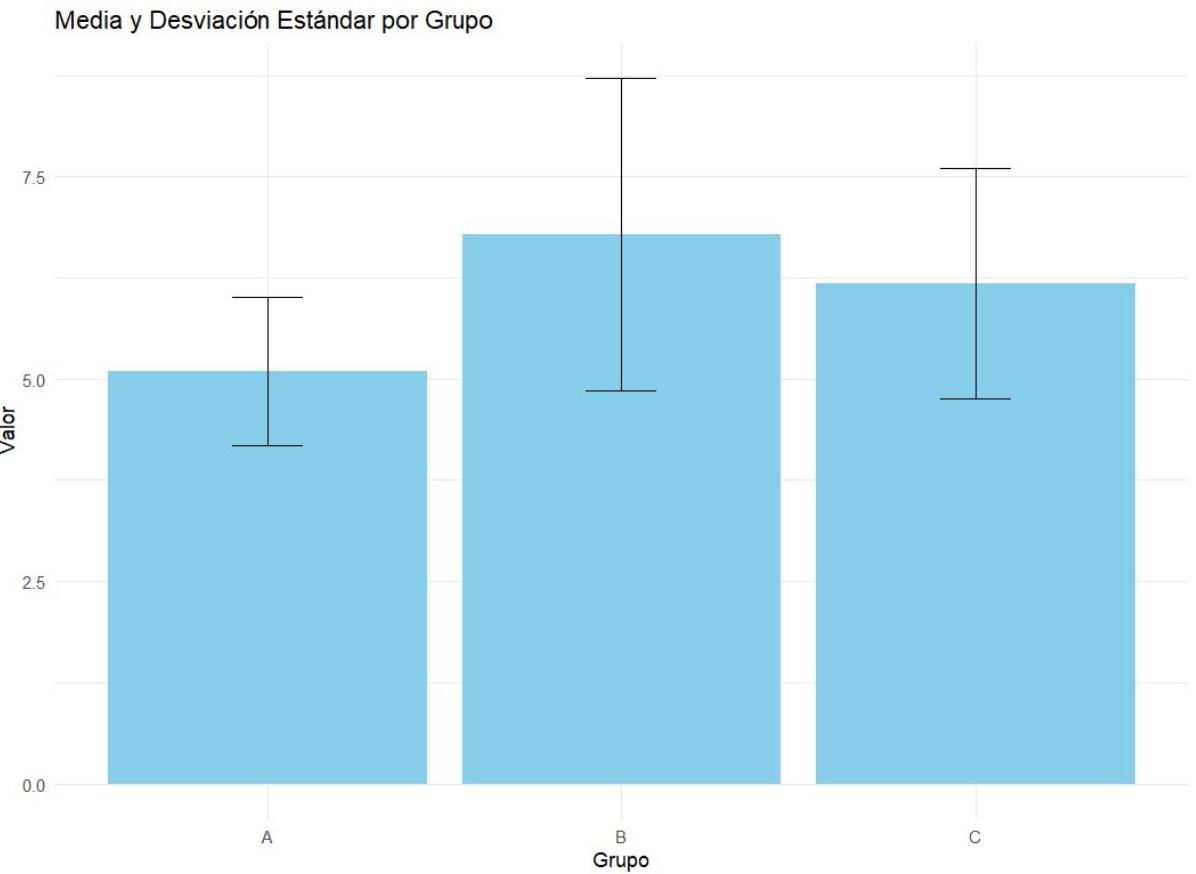
$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

donde:

- $N$  es el tamaño de la muestra.
- $x_i$  es cada observación individual,
- $\bar{x}$  es la media de la muestra.

# Standard deviation

$$S = \sqrt{S^2}$$



## **COEFICIENTE DE VARIACIÓN (Variabilidad Relativa)**

$$CV = \frac{S}{\bar{x}} * 100\%$$

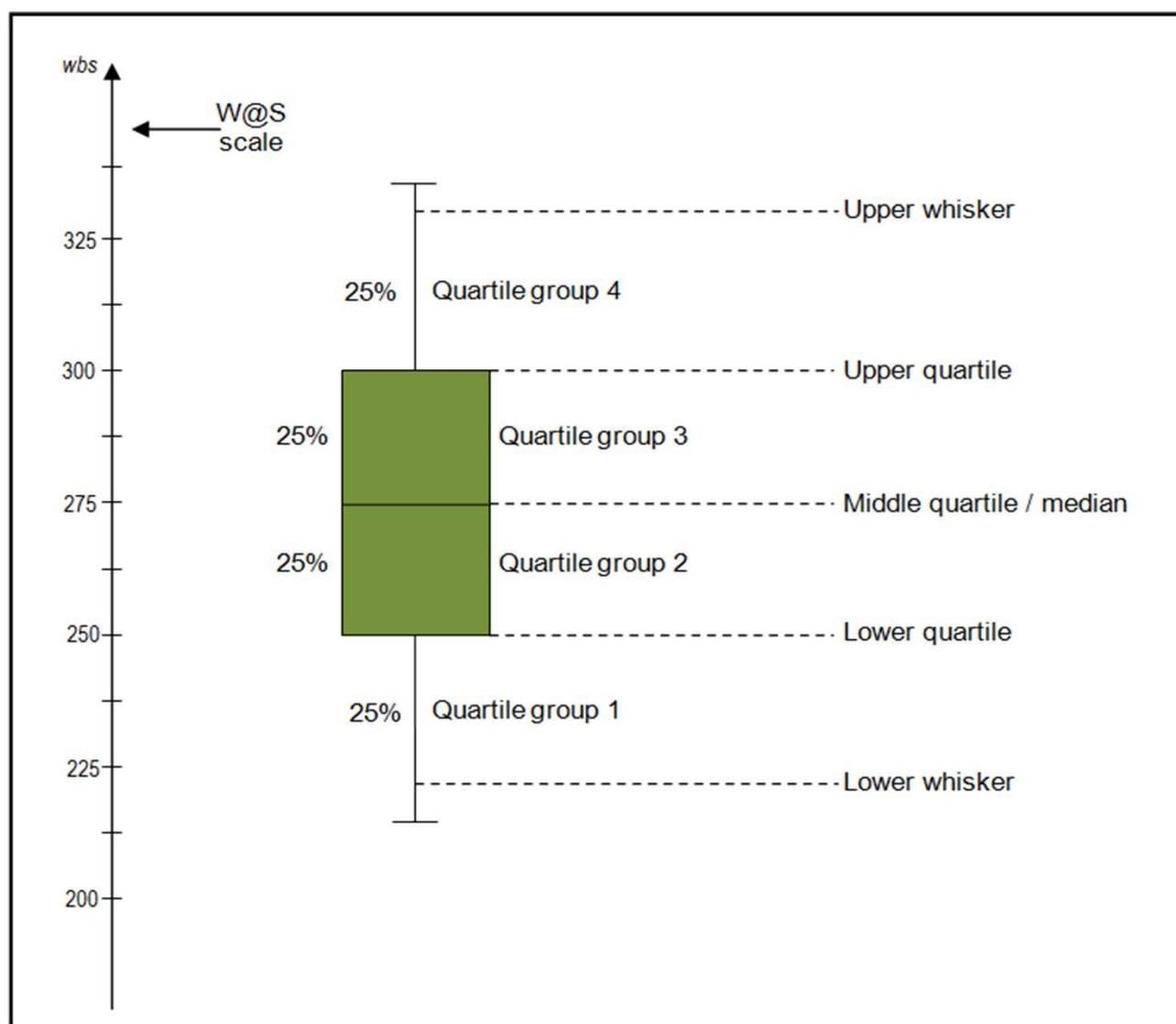
- ✓ Elimina los problemas de la escala de medida, permitiendo comparar entre variables diferentes.
- ✓ Su cercanía a 0 implica homogeneidad de la característica, valores altos corresponden a características heterogéneas.
- ✗ No debe usarse cuando la variable presenta valores negativos o en escalas de intervalos

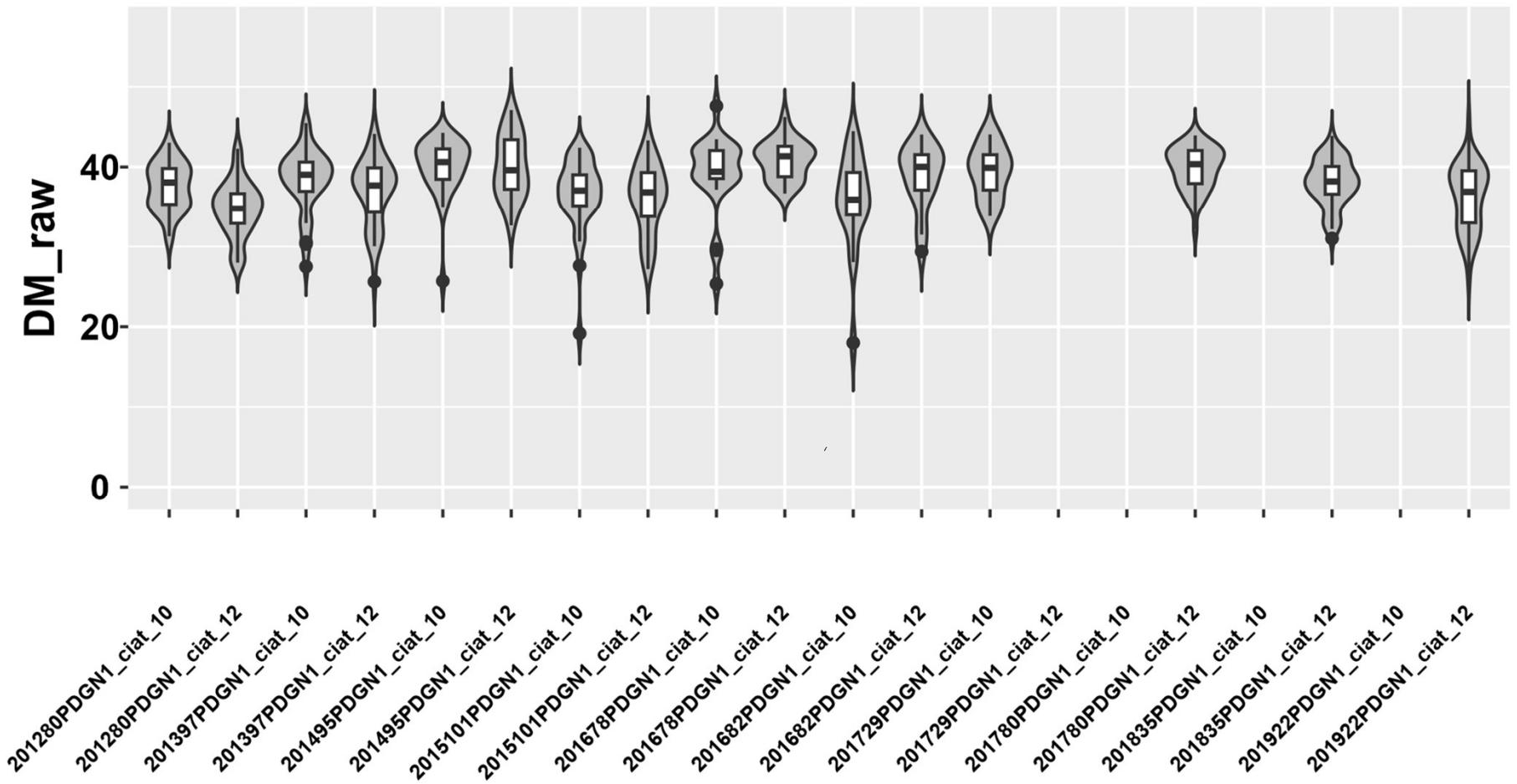
# Indicadores de Posición

Permiten hacerse una idea acerca de la forma de la distribución una variable y su dispersión. En ocasiones sirven como criterio de selección.

**Cuartiles:** Dividen la distribución en cuatro partes porcentualmente iguales, Q1, Q2, Q3

- Q1= Valor que es superior al 25% de las observaciones
- Q2= Valor que es superior al 50% de las observaciones
- Q3 = Valor que es superior al 75% de las observaciones





**Percentiles:**

P<sub>1</sub>, P<sub>2</sub>, P<sub>3</sub>, ..., P<sub>100</sub>

**Deciles:**

D<sub>1</sub>, D<sub>2</sub>, D<sub>3</sub>, ...D<sub>9</sub>, D<sub>10</sub>

**Cuartiles:**

Q<sub>1</sub>, Q<sub>2</sub>, Q<sub>3</sub>

Nótese que tanto los deciles como los cuartiles se pueden representar en función de los percentiles:

P<sub>25</sub> = Q<sub>1</sub>

P<sub>50</sub> = Q<sub>2</sub> = Me

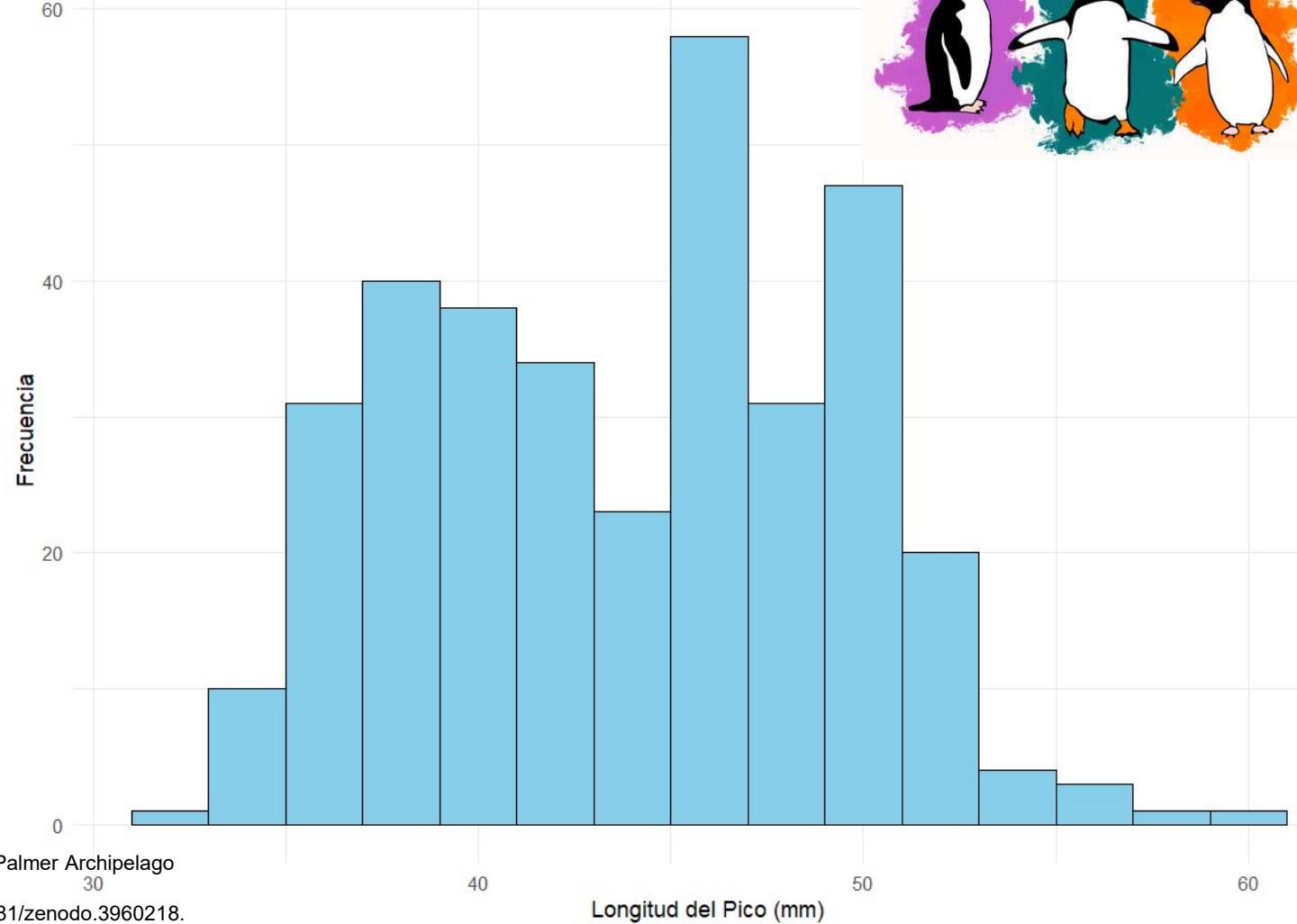
P<sub>75</sub> = Q<sub>3</sub>

P<sub>10</sub> = D<sub>1</sub>, P<sub>20</sub> = D<sub>2</sub>, P<sub>30</sub> = D<sub>3</sub>, ..., P<sub>50</sub> = Me, ..., P<sub>90</sub> = D<sub>9</sub>, P<sub>100</sub> = D<sub>10</sub>

*"La representación gráfica de los Cuartiles es el diagrama de cajas y alambres"*

# Histogramas

Distribución de la Longitud del Pico de los Pingüinos

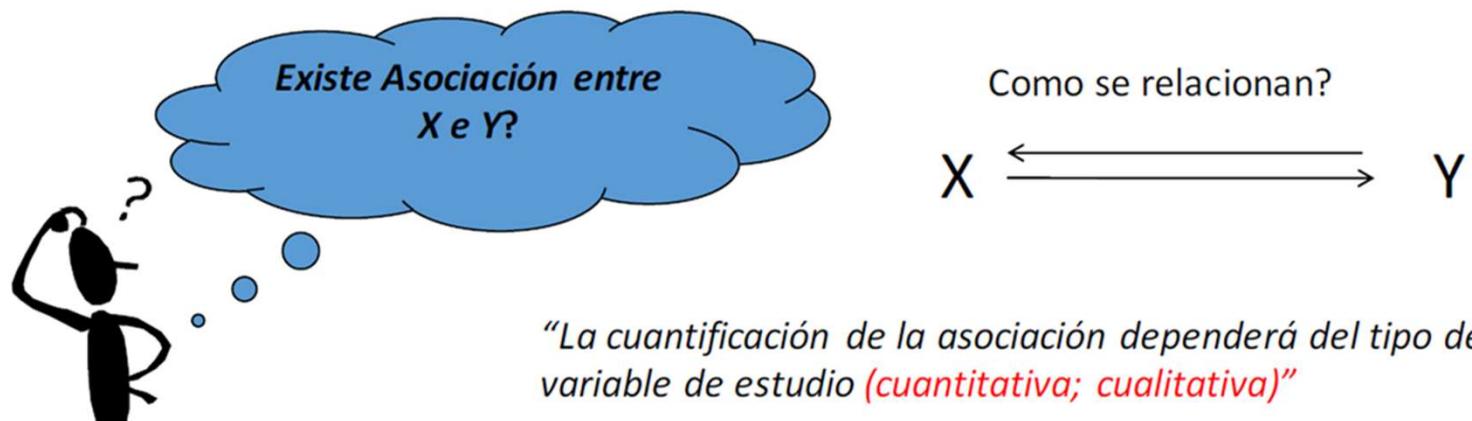


Horst AM, Hill AP, Gorman KB (2020). palmerpenguins: Palmer Archipelago (Antarctica) penguin data. R package version 0.1.0.  
<https://allisonhorst.github.io/palmerpenguins/>. doi: 10.5281/zenodo.3960218.

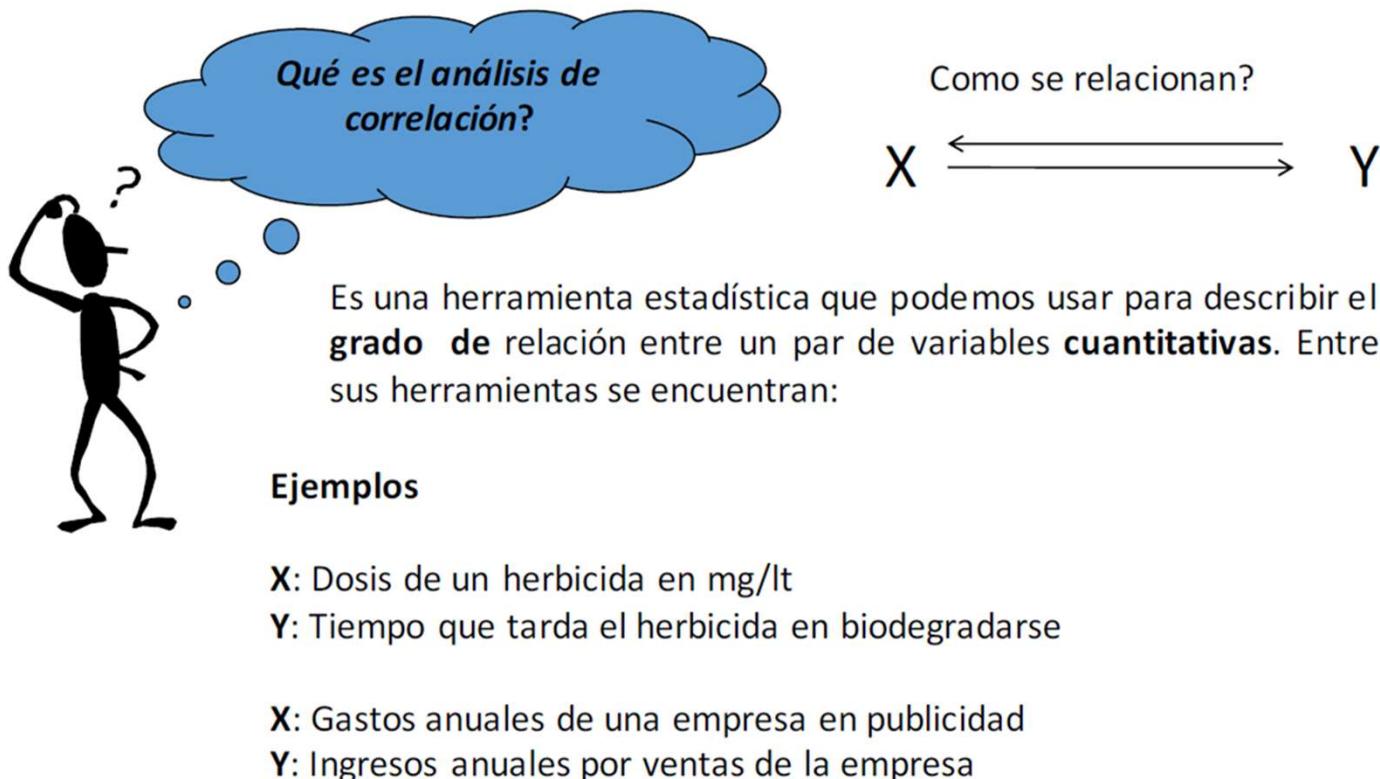
# Medida de Asociación

**Definición:** Medición que intenta cuantificar la relación estadística existente entre dos características definidas “X” e “Y”.

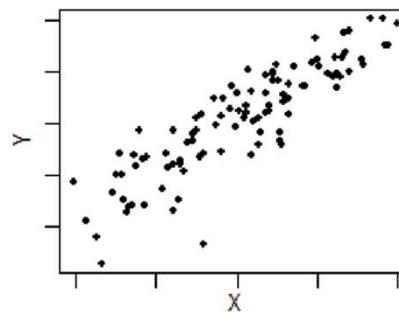
*“Las relaciones encontradas no son necesariamente del tipo causal, son solo relaciones estadísticas”*



# Análisis de Correlación



- I. *Gráficos de Dispersión*: representación gráfica simultánea del par de variables estudiadas.



- I. *Coeficientes de Correlación (r)*: indicador que mide la fuerza de la relación entre un par de variables.

$$r = \frac{Cov(x, y)}{\sqrt{V(x)V(y)}} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{\left[n \sum (x^2) - (\sum x)^2\right] \left[n \sum (y^2) - (\sum y)^2\right]}}$$



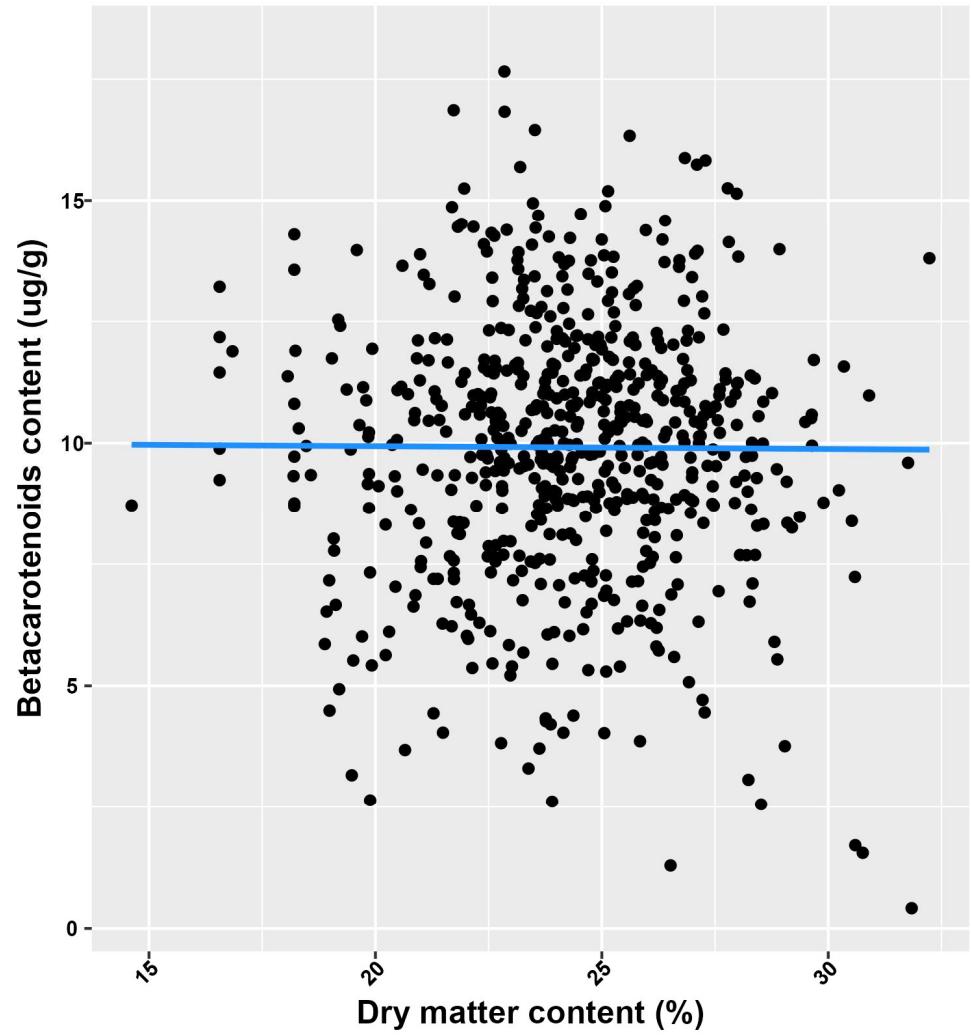
Karl Pearson

- I. *Ajuste de Modelos Explicativos*: búsqueda de una expresión matemática que permita predecir, de forma aproximada, el valor de Y a través del conocimiento del valor de X.

$$Y = a + bX$$

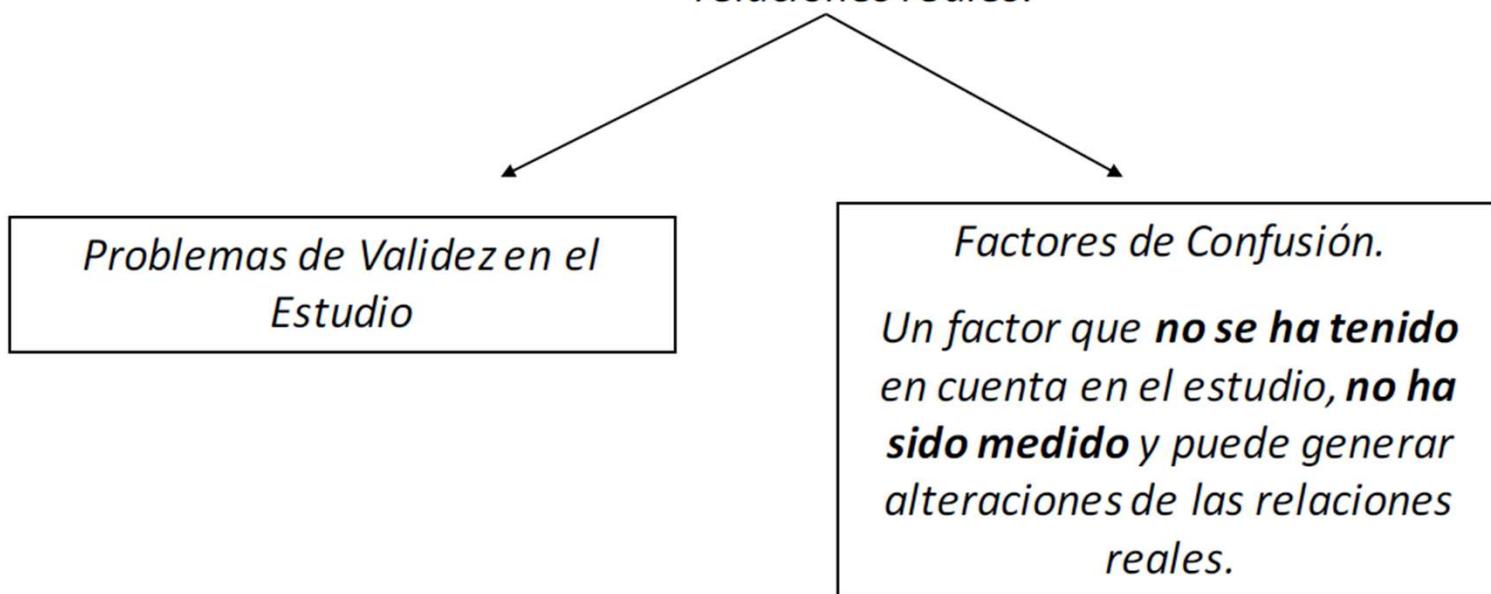
# Scatter plot

Plot a corplot with Good corr

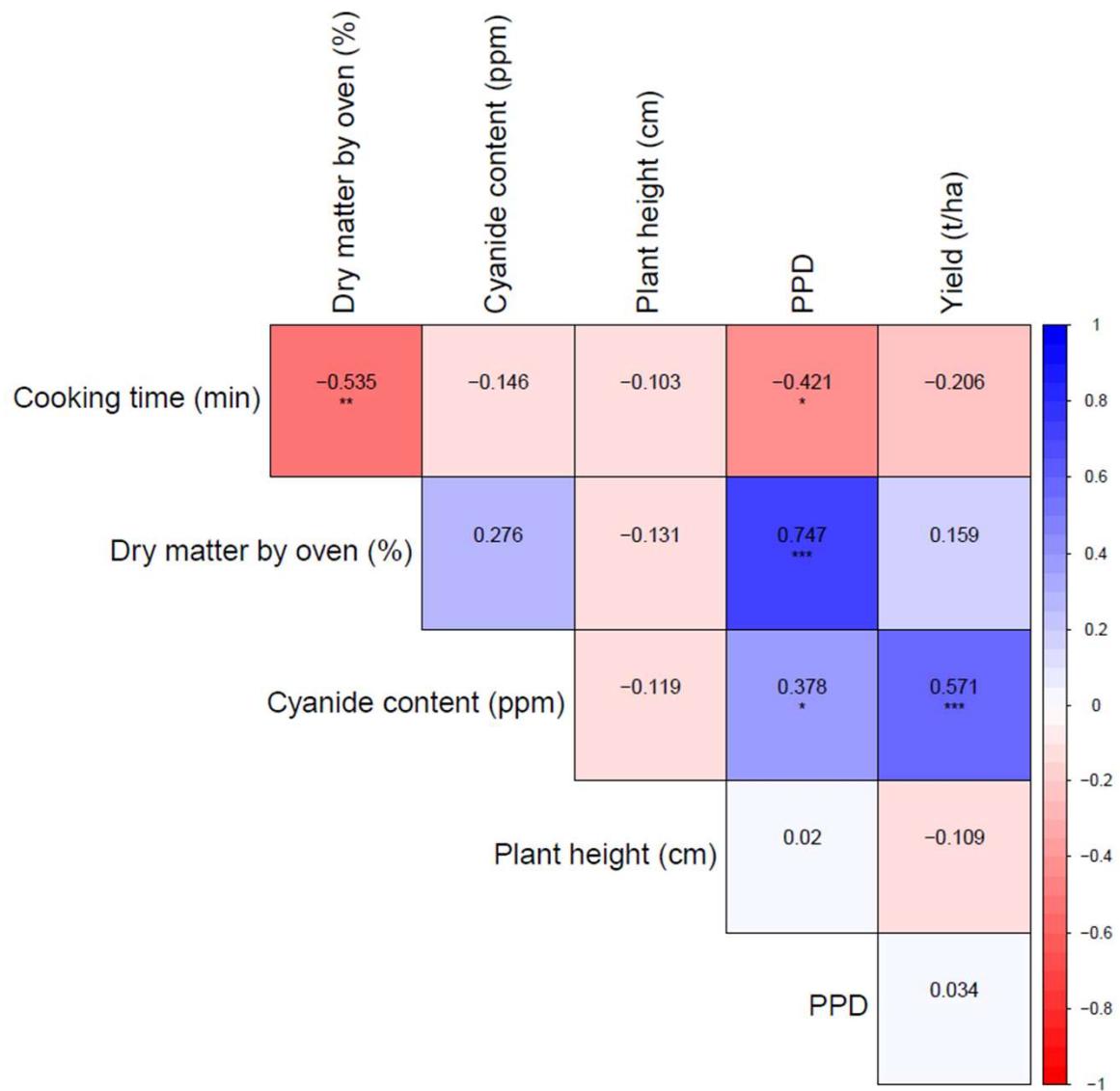


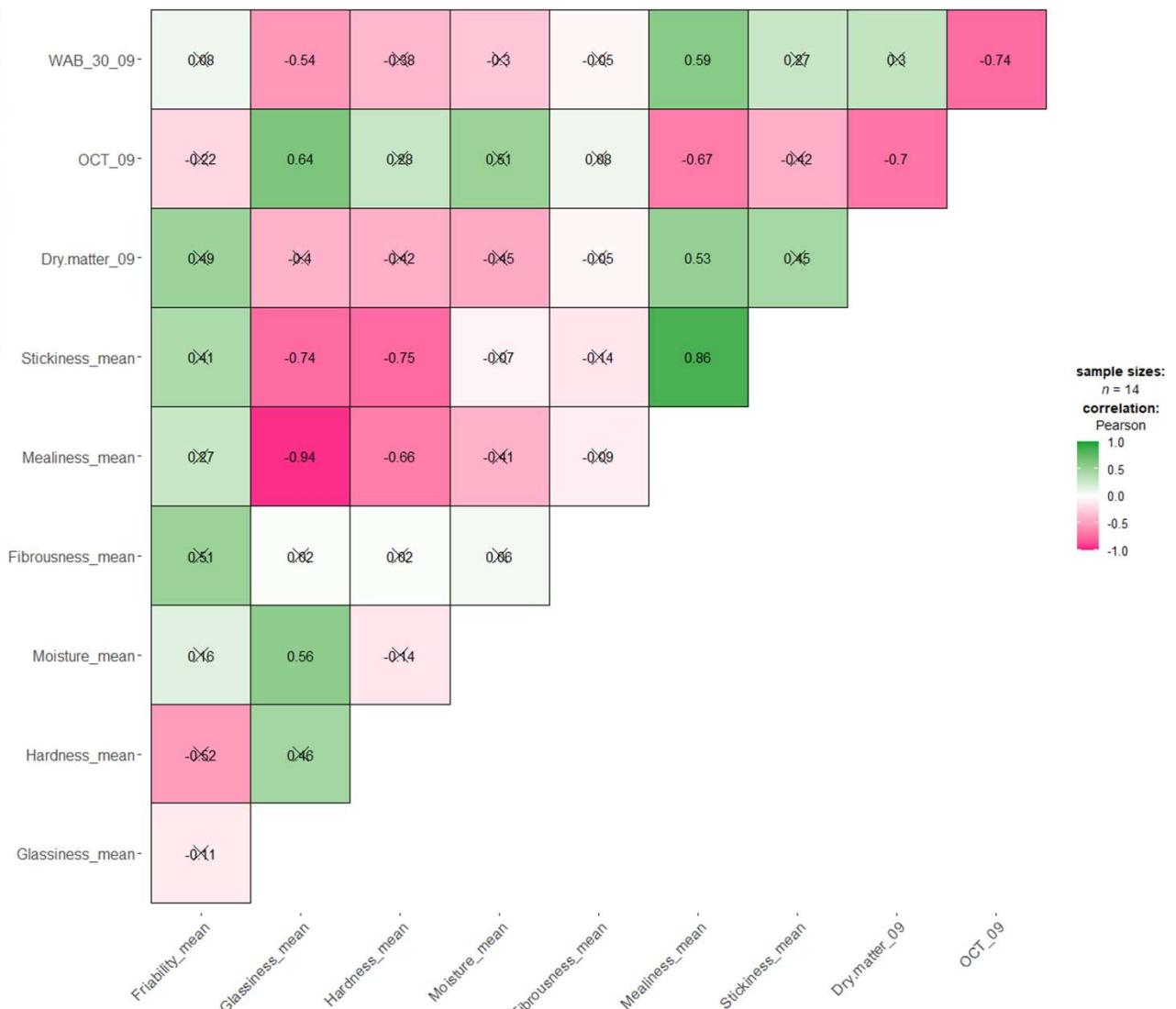
# Precauciones en las correlaciones

*En la práctica es muy común encontrar relaciones estadísticas donde no existen relaciones reales.*

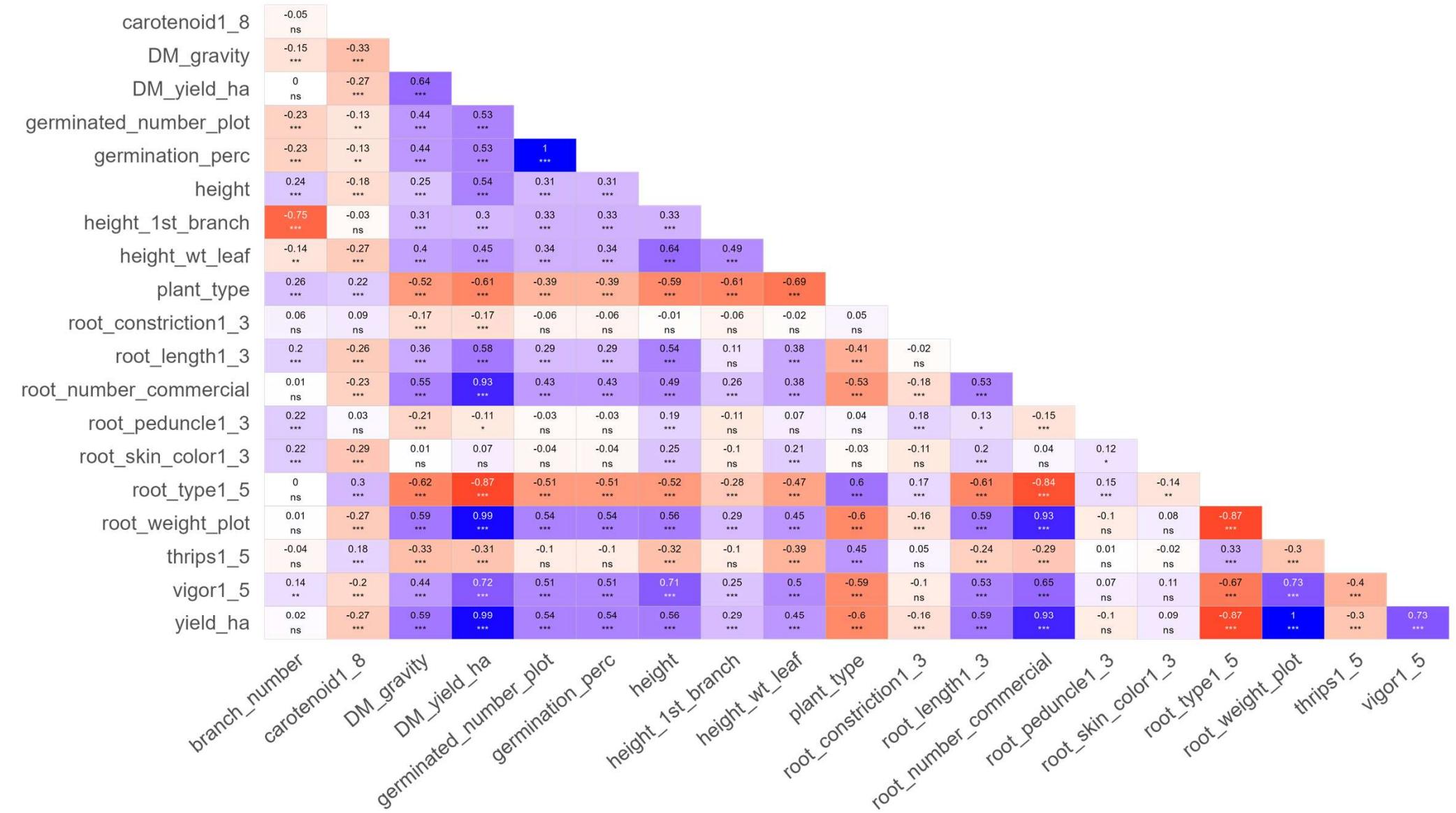








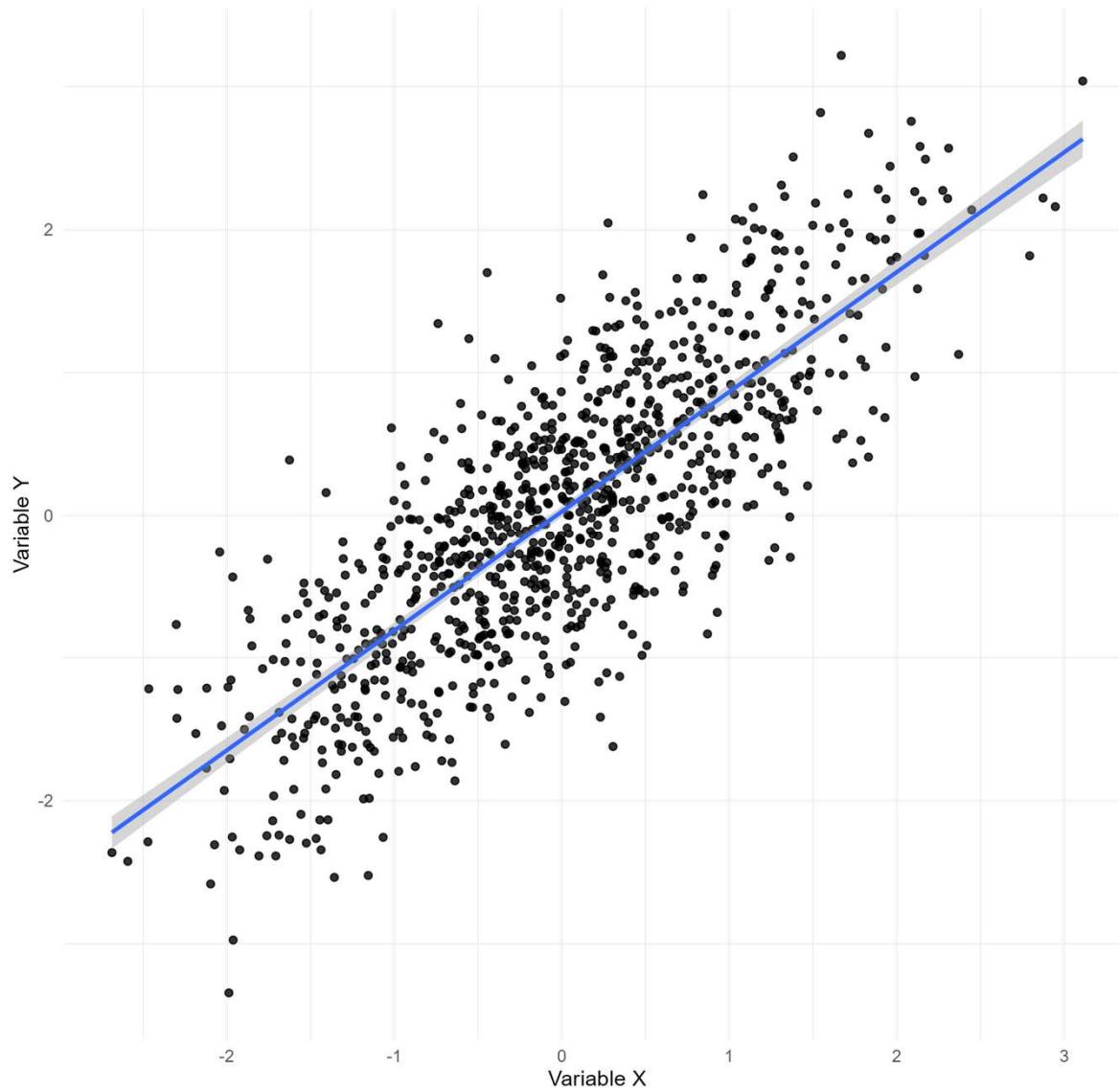
X = non-significant at  $p < 0.05$  (Adjustment: None)



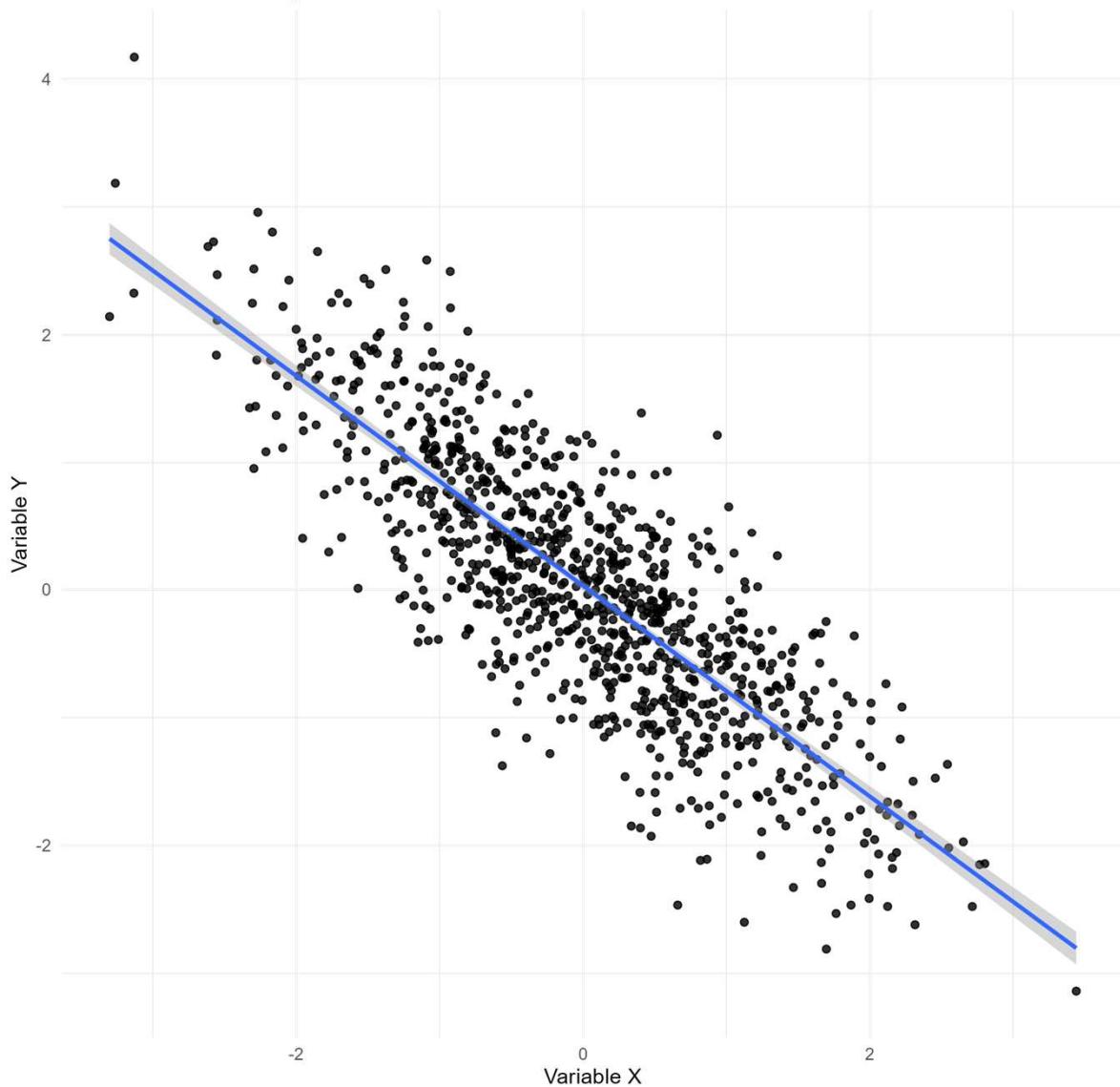
¿Como valorar la  
correlación?



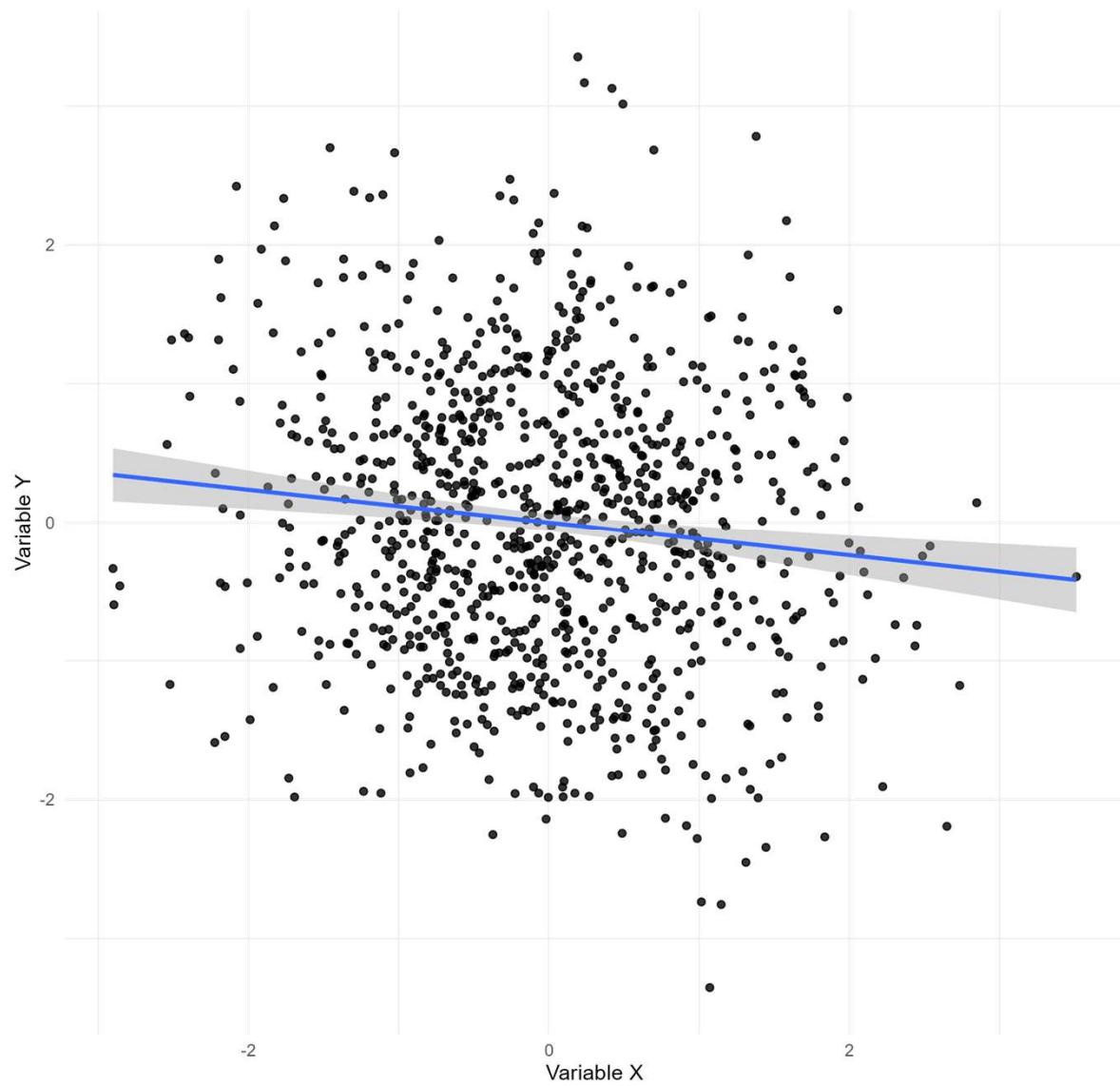
Scatter Plot with Positive Linear Correlation



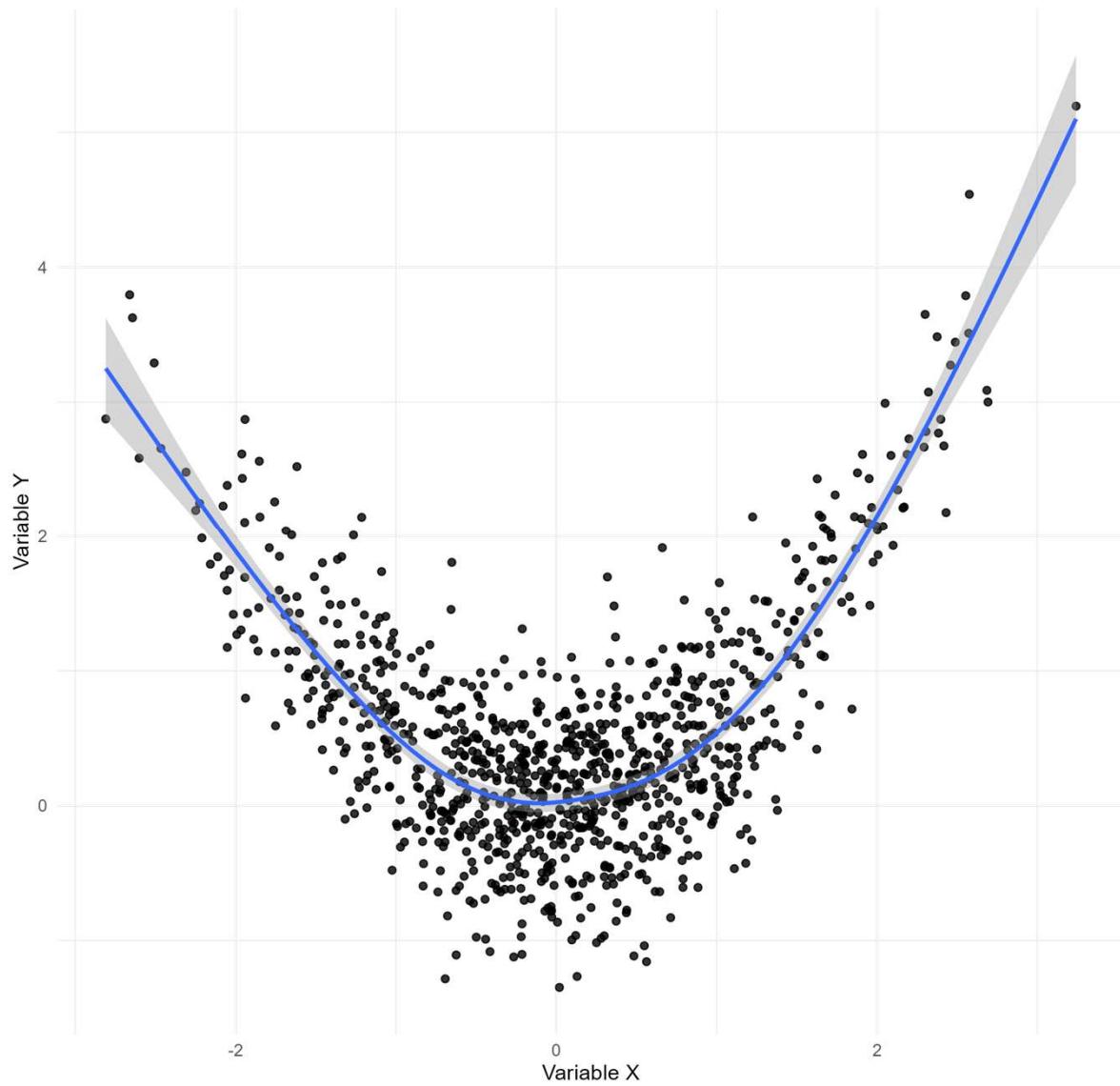
Scatter Plot with Negative Linear Correlation



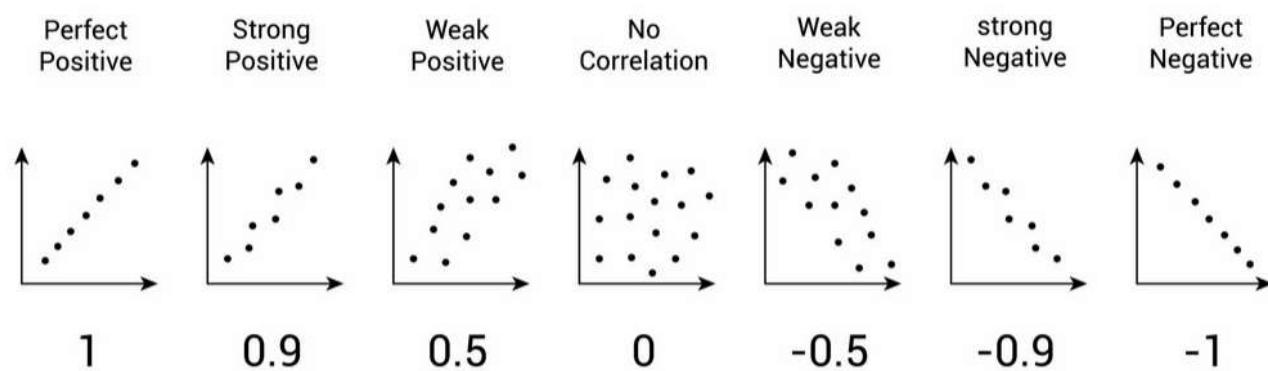
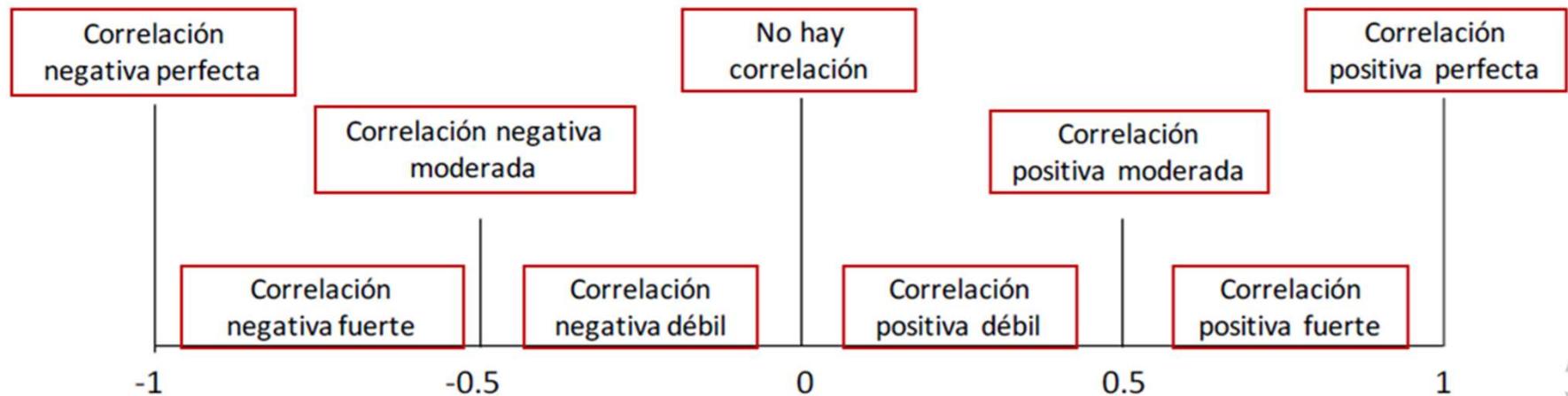
Scatter Plot with non Linear Correlation

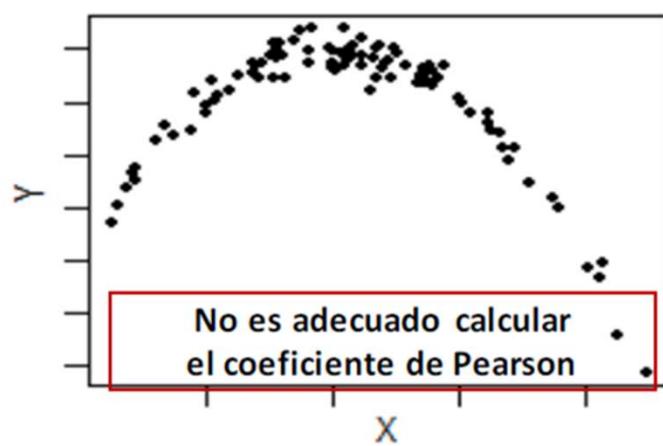
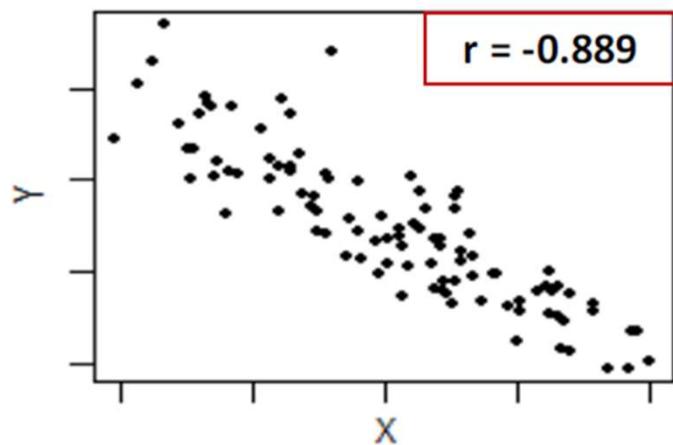
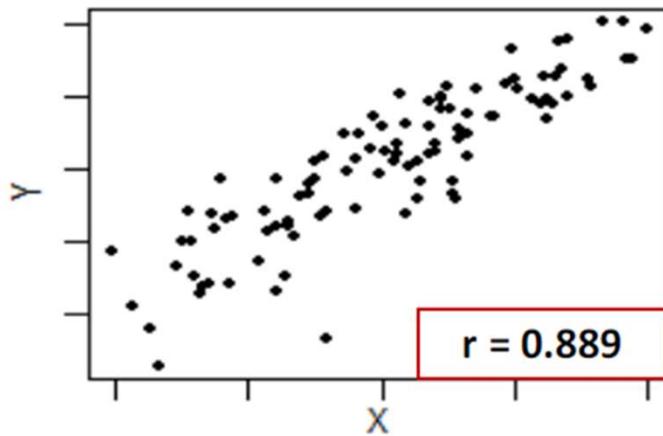
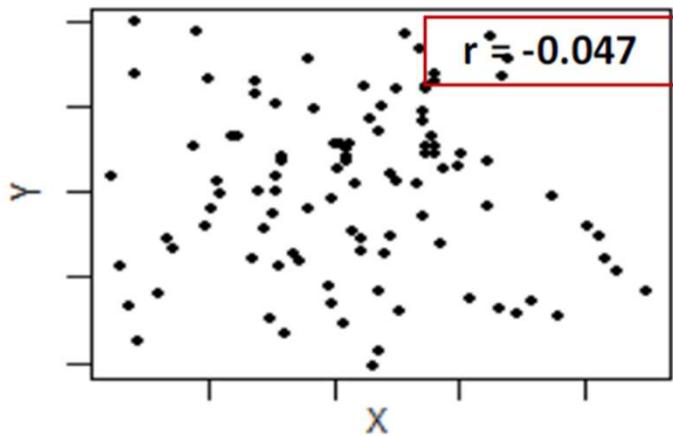


Scatter Plot with Non-Linear Correlation



# Propiedades de $r$





# Análisis Exploratorio de Datos



Enfrentado las situaciones

# Antes de Continuar.....

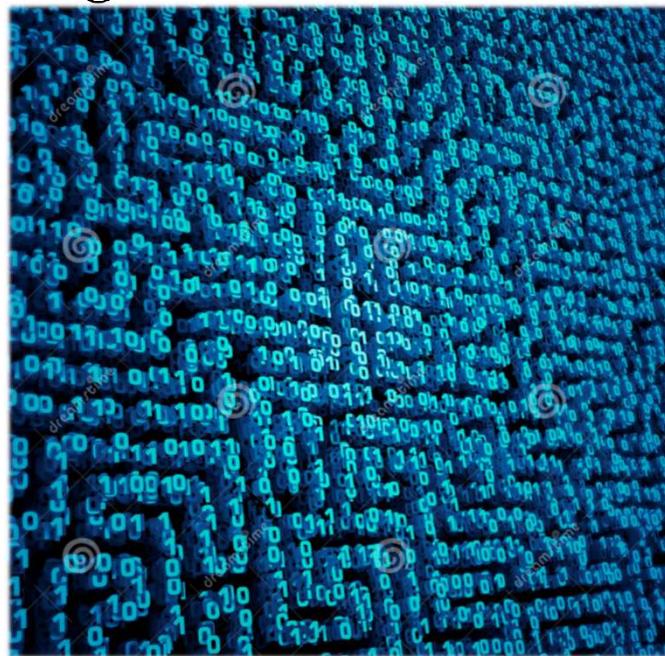
Todo estudio basado en datos, sin importar su alcance, debe superar la fase inicial del análisis exploratorio.

“Tabular, graficar, resumir, para identificar patrones y comportamientos regulares y presencia de irregularidades en los datos”

Existen patrones de comportamiento regular en los datos?

- Se presentan datos atípicos? Que hacer con ellos?
- Como se relacionan las variables de análisis?
- Existen diferencias en el comportamiento de la variable entre grupos de análisis?

El Análisis Exploratorio de datos no es una rutina, es una actividad individual en la cual el analista escoge su ruta.



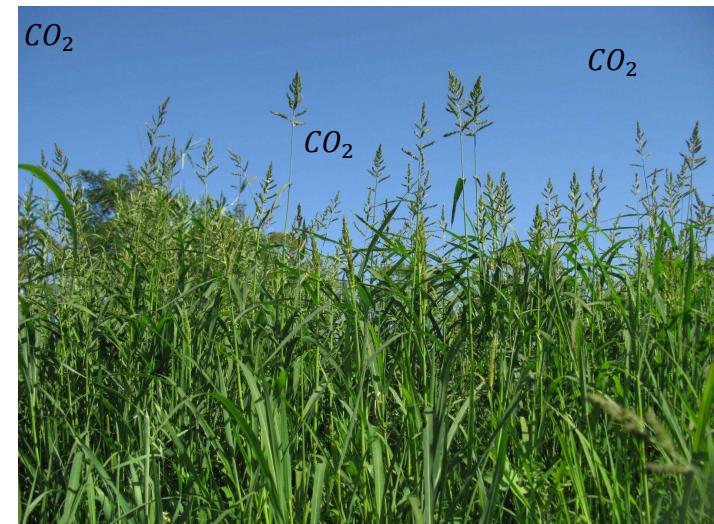
Para este tipo de análisis no existe una receta, existen herramientas, cuya implementación dependerá **de la tipología de variables de análisis** y de la necesidad de síntesis de la información.

# Tarea

Llevar a cabo el análisis exploratorio de datos del siguiente dataset:

# Carbon Dioxide Uptake in Grass Plants

The CO<sub>2</sub> data frame has 84 rows and 5 columns of data from an experiment on the cold tolerance of the grass species *Echinochloa crus-galli*.



## **Plant**

- An ordered factor with levels  $Qn1 < Qn2 < Qn3 < \dots < Mc1$  giving a unique identifier for each plant.

## **Type**

- A factor with levels Quebec Mississippi giving the origin of the plant

## **Treatment**

- A factor with levels nonchilled chilled

## **conc**

- A numeric vector of ambient carbon dioxide concentrations (mL/L).

## **uptake**

- A numeric vector of carbon dioxide uptake rates  $\left( \frac{\mu mol}{m^2 sec} \right)$

Plant	Type	Treatment	conc	uptake
Qn1	Quebec	nonchilled	95	16
Qn1	Quebec	nonchilled	175	30.4
Qn1	Quebec	nonchilled	250	34.8
Qn1	Quebec	nonchilled	350	37.2
Qn1	Quebec	nonchilled	500	35.3
Qn1	Quebec	nonchilled	675	39.2
Qn1	Quebec	nonchilled	1000	39.7
Qn2	Quebec	nonchilled	95	13.6
Qn2	Quebec	nonchilled	175	27.3
Qn2	Quebec	nonchilled	250	37.1
Qn2	Quebec	nonchilled	350	41.8
Qn2	Quebec	nonchilled	500	40.6
Qn2	Quebec	nonchilled	675	41.4
Qn2	Quebec	nonchilled	1000	44.3
Qn3	Quebec	nonchilled	95	16.2
Qn3	Quebec	nonchilled	175	32.4
Qn3	Quebec	nonchilled	250	40.3
Qn3	Quebec	nonchilled	350	42.1
Qn3	Quebec	nonchilled	500	42.9
Qn3	Quebec	nonchilled	675	43.9
Qn3	Quebec	nonchilled	1000	45.5
Qc1	Quebec	chilled	95	14.2
Qc1	Quebec	chilled	175	24.1
Qc1	Quebec	chilled	250	30.3
Qc1	Quebec	chilled	350	34.6
Qc1	Quebec	chilled	500	32.5
Qc1	Quebec	chilled	675	35.4
Qc1	Quebec	chilled	1000	38.7
Qc2	Quebec	chilled	95	9.3
Qc2	Quebec	chilled	175	27.3
Qc2	Quebec	chilled	250	35
Qc2	Quebec	chilled	350	38.8
Qc2	Quebec	chilled	500	38.6
Qc2	Quebec	chilled	675	37.5
Qc3	Quebec	chilled	1000	42.4
Qc3	Quebec	chilled	95	15.1
Qc3	Quebec	chilled	175	21
Qc3	Quebec	chilled	250	38.1
Qc3	Quebec	chilled	350	34
Qc3	Quebec	chilled	500	38.3
Qc3	Quebec	chilled	675	39.3
Qc3	Quebec	chilled	1000	41.4
Mn1	Mississippi	nonchilled	95	10.6
Mn1	Mississippi	nonchilled	175	19.2
Mn1	Mississippi	nonchilled	250	26.2
Mn1	Mississippi	nonchilled	350	30
Mn1	Mississippi	nonchilled	500	30.9
Mn1	Mississippi	nonchilled	675	32.4
Mn1	Mississippi	nonchilled	1000	35.5
Mn2	Mississippi	nonchilled	95	12
Mn2	Mississippi	nonchilled	175	22
Mn2	Mississippi	nonchilled	250	30.6
Mn2	Mississippi	nonchilled	350	31.8
Mn2	Mississippi	nonchilled	500	32.4
Mn2	Mississippi	nonchilled	675	31.1
Mn2	Mississippi	nonchilled	1000	31.5
Mn3	Mississippi	nonchilled	95	11.3
Mn3	Mississippi	nonchilled	175	19.4
Mn3	Mississippi	nonchilled	250	25.8
Mn3	Mississippi	nonchilled	350	27.9
Mn3	Mississippi	nonchilled	500	28.5
Mn3	Mississippi	nonchilled	675	28.1
Mn3	Mississippi	nonchilled	1000	27.6
Mc1	Mississippi	chilled	95	10.5
Mc1	Mississippi	chilled	175	14.9
Mc1	Mississippi	chilled	250	18.1
Mc1	Mississippi	chilled	350	18.9
Mc1	Mississippi	chilled	500	19.5
Mc1	Mississippi	chilled	675	22.2
Mc1	Mississippi	chilled	1000	21.9
Mc2	Mississippi	chilled	95	7.7
Mc2	Mississippi	chilled	175	11.4
Mc2	Mississippi	chilled	250	12.3
Mc2	Mississippi	chilled	350	13
Mc2	Mississippi	chilled	500	12.5
Mc2	Mississippi	chilled	675	13.7
Mc2	Mississippi	chilled	1000	14.4
Mc3	Mississippi	chilled	95	10.6
Mc3	Mississippi	chilled	175	18
Mc3	Mississippi	chilled	250	17.9
Mc3	Mississippi	chilled	350	17.9
Mc3	Mississippi	chilled	500	17.9
Mc3	Mississippi	chilled	675	18.9
Mc3	Mississippi	chilled	1000	19.9