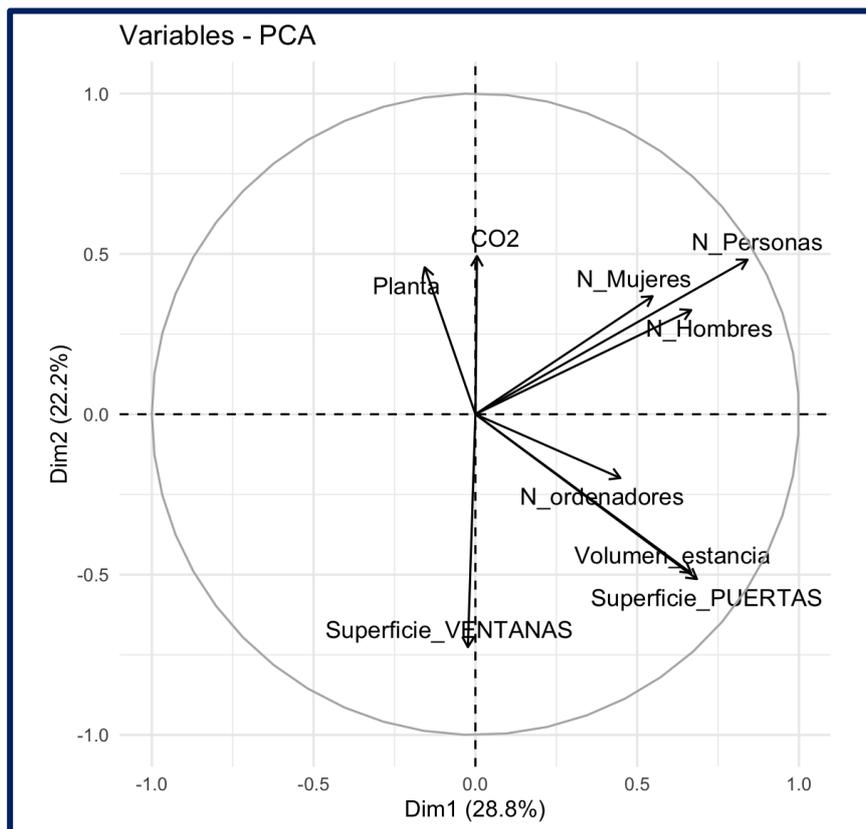


6.- ANÁLISIS DE COMPONENTES PRINCIPALES

La representación bidimensional de todas las variables por parejas, si el número de variables es muy alto, sería inviable. Además, como hemos visto anteriormente, solo representan una fracción de la información contenida en los datos.

El análisis de componentes principales es una estrategia adecuada cuando se examinan muchas variables y se quiere ver las relaciones que hay entre ellas. Se basa en un análisis en varias dimensiones donde las variables se proyectan sobre los ejes de dichas dimensiones, de forma que las variables que quedan sobre el mismo eje contribuyen positiva o negativamente dependiendo del sentido en el que estén en el eje.

En este gráfico se representan las variables cuantitativas estudiados en dos dimensiones, en cada una de las dimensiones se indica el % de varianza explicado por el primer componente (Dim1) y el segundo componente (Dim2). Las variables positivamente correlacionadas se agrupan y aparecerán juntas o muy cercanas. Las variables negativamente correlacionadas aparecerán al lado opuesto del origen o en cuadrantes diferentes. La longitud de las variables mide la calidad de la representación de la variable, cuanto más cerca está de la circunferencia de correlación representada, más calidad tiene su representación, es decir mayor variabilidad de esta variable está contenida en la representación.



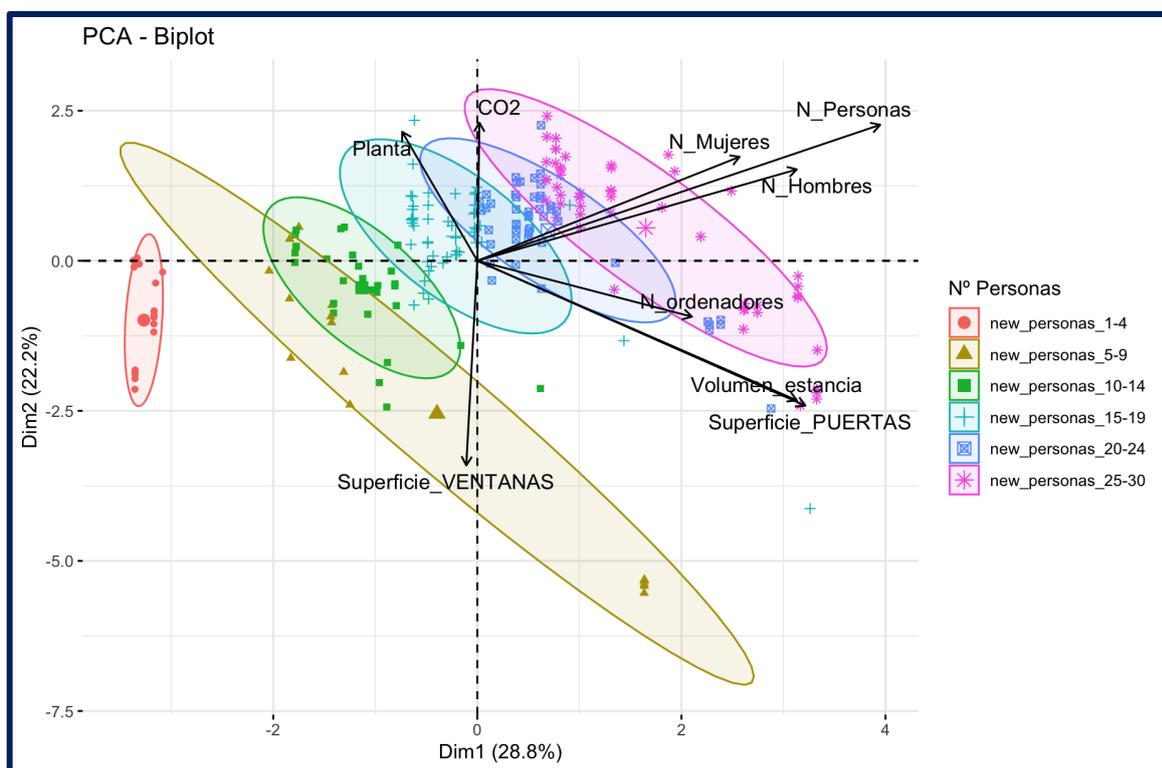
Gráfica 81: Representación vectorial de variables estudiadas y círculo de correlación

Variable	correlation	p.value
CO2	4.93113e-01	8.05729e-15
N_Personas	4.82139e-01	3.77859e-14
Planta	4.58091e-01	9.26255e-13
N_Mujeres	3.68007e-01	1.98767e-08
N_Hombres	3.24688e-01	9.05412e-07
Nordenadores	-1.98915e-01	3.11228e-03
Volumen_estancia	-4.95231e-01	5.94198e-15
Superficie_PUERTAS	-5.13812e-01	3.74215e-16
Superficie_VENTANAS	-7.26050e-01	3.78560e-37

Tabla 33: Coeficientes de correlación de las variables analizadas.

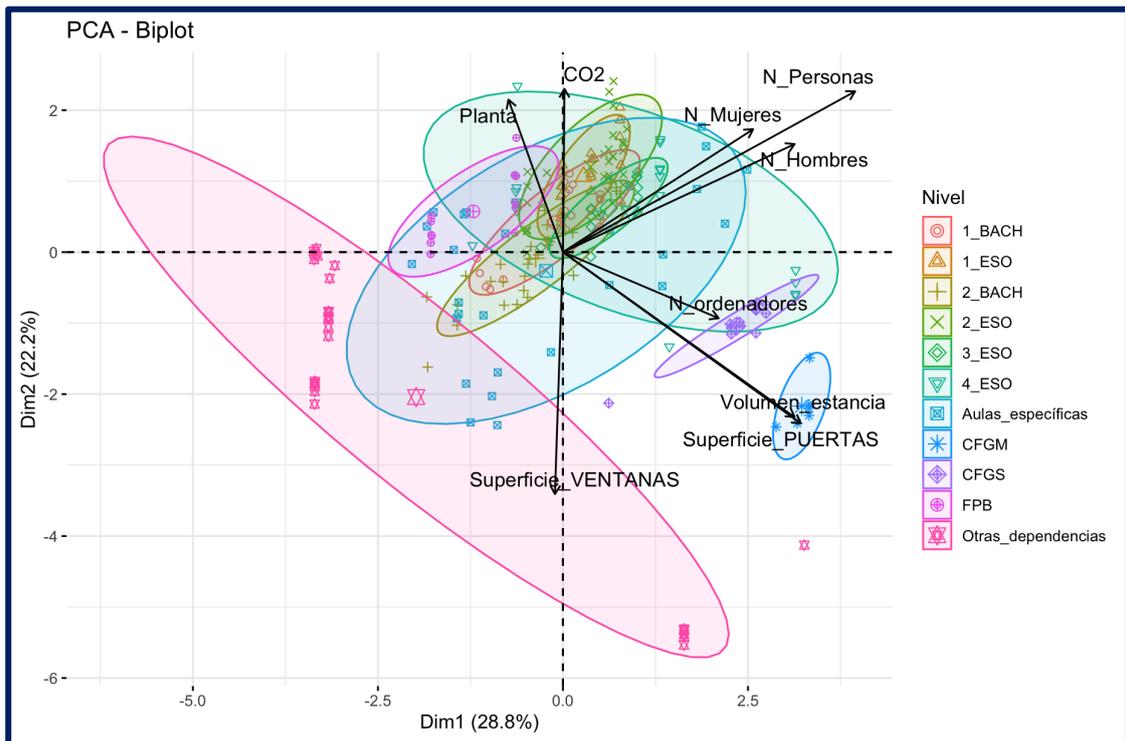
En la Gráfica 81 se observa que las variables CO₂ y planta están situadas sobre el eje de dimensión 1 en sentido positivo, mientras que la variable Superficie_Ventanas esta sobre ese mismo eje, pero en sentido contrario. Eso quiere decir que al incrementarse la planta se incrementa el CO₂, y al aumentar la superficie de ventanas disminuye el CO₂. Lo mismo ocurre con las variables volumen estancia y superficie PUERTAS, que también están orientadas en sentido contrario a planta y CO₂, y en sentido negativo, por lo que al incrementarse el volumen y la superficie de puertas disminuye el CO₂. Por el contrario, las variables Número de Personas, independientemente de que sean hombres o mujeres están en el lado positivo y cerca de la variable CO₂, por lo que al incrementarse el número de personas se incrementa el CO₂.

En nuestro PCA hemos estudiado varias variables cuantitativas que están representadas por los vectores, pero para poder interpretar mejor los datos podemos representar sobre esta misma gráfica las variables cualitativas y analizar cómo se agrupan los datos respecto de estas variables cualitativas, proporcionando importante información en relación con las cuantitativas.



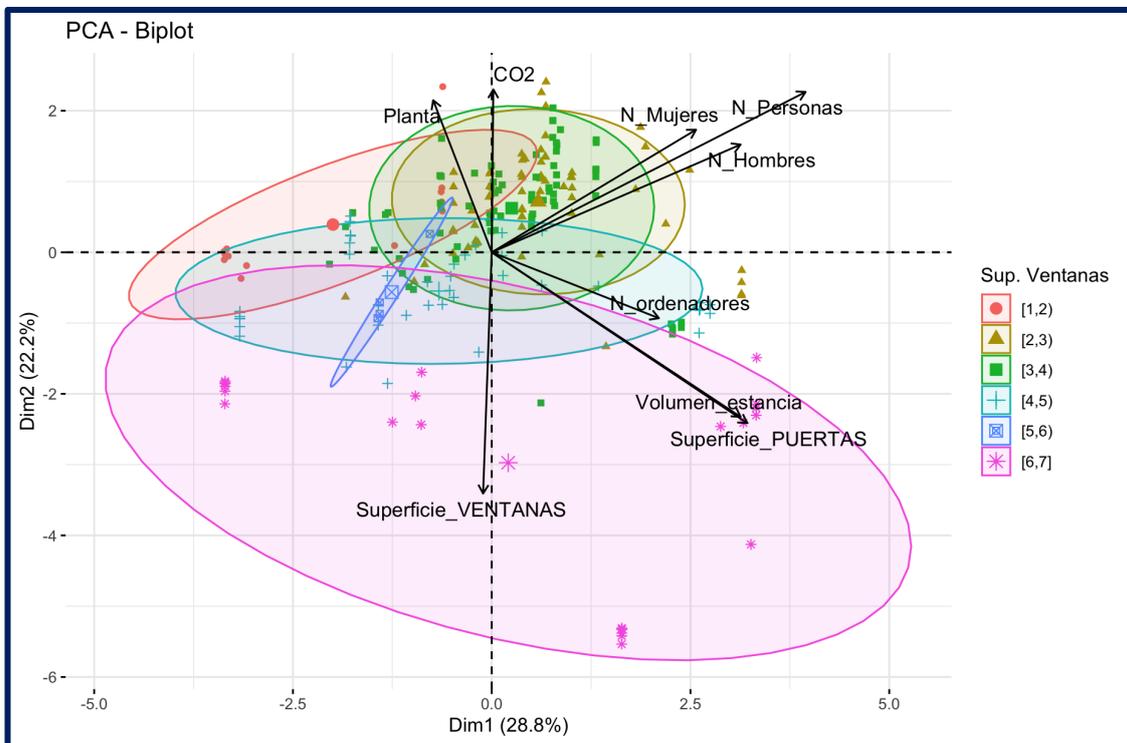
Gráfica 82: Gráfico Biplot representando agrupamiento por número de personas.

Respecto al agrupamiento de la variable número de personas representado en la Gráfica 82 se evidencia de forma muy clara cómo los grupos se alinean de forma paralela al vector número de personas. Las elipses de colores son los rangos establecidos en la variable cualitativa "new_personas" en la que los datos se agrupan dentro de intervalos que establecen un rango de números de personas. Así el grupo de 1-4 personas, marcado en color rojo, aparece distanciado y lejos del vector, lo que significa que contribuyen poco a la acumulación de CO₂ en la estancia, mientras los grupos con elevado número de personas contribuyen con elevadas cantidades de CO₂. Claramente se ilustra el efecto de esta variable, a mayor número de personas mayor CO₂ acumulado en el ambiente.



Gráfica 83: Gráfico Biplot representando agrupamiento por niveles

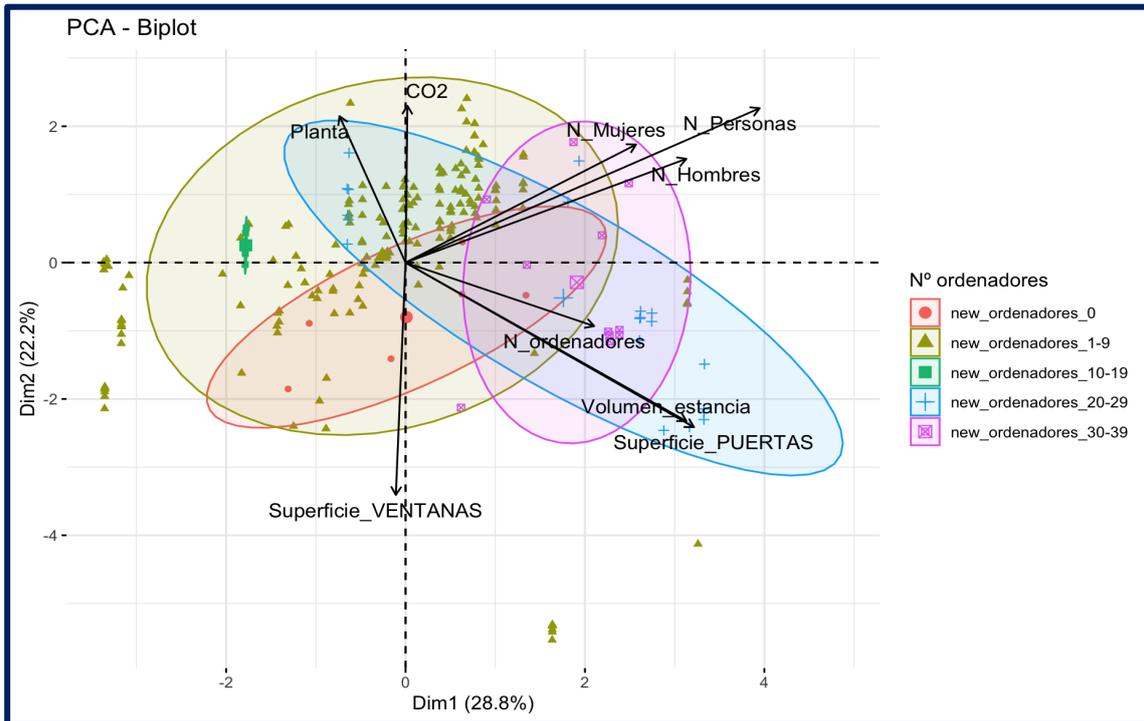
Con respecto a los agrupamientos por niveles (Gráfica 83), la distribución de los datos es más irregular, pero observamos cómo las elipses de 1º de ESO y 2º de ESO están muy juntas y pegadas al vector, lo que indica mayor CO₂ acumulado en estos niveles. Por el contrario, las elipses de los ciclos y otras dependencias aparecen más alejadas del vector.



Gráfica 84: Gráfico Biplot representando agrupamiento por superficie ventanas.

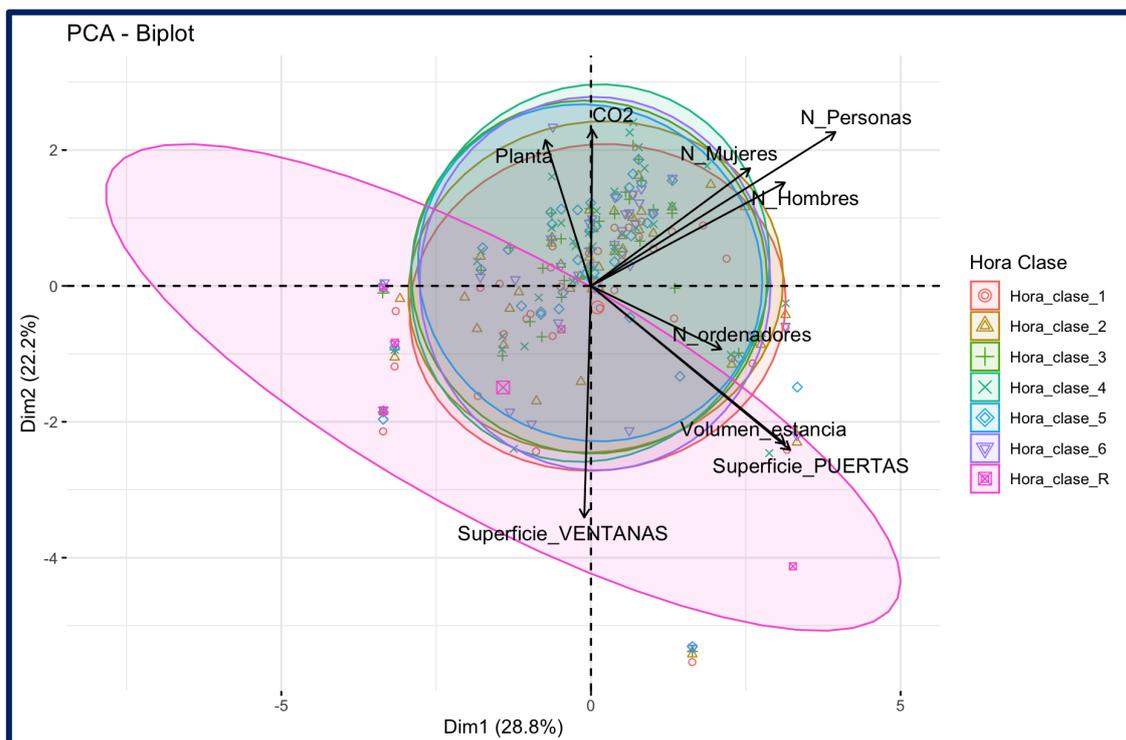
En la Gráfica 84 observamos que las estancias con mayor superficie de ventanas acumulan muy poco CO₂. Los espacios con 6-7 m² de área de entrada de aire están representados por una gran elipse rosa que ocupa un espacio extenso y se encuentra lejos del origen de coordenadas y muy lejos del

extremo del vector CO₂. Se corrobora así que la correlación entre CO₂ y superficie de ventanas es negativa.



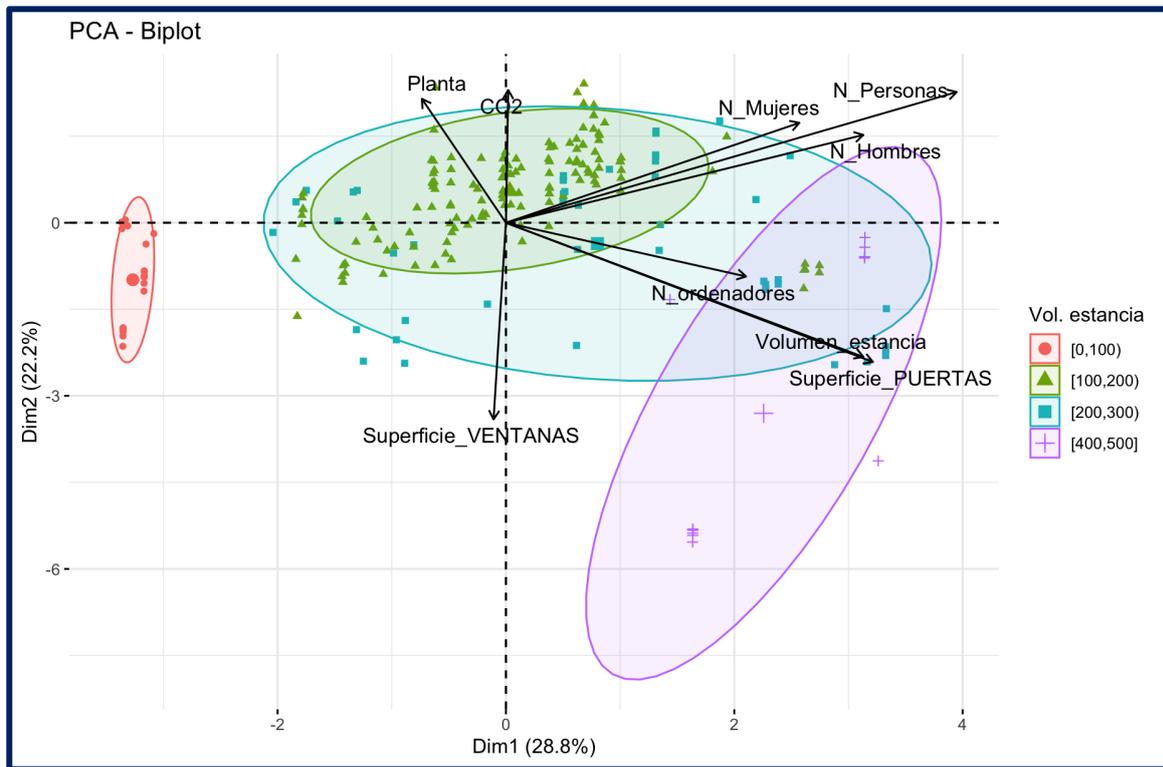
Gráfica 85: Gráfico Biplot con agrupamiento por número de ordenadores.

En la Gráfica 85 la distribución de las elipses no guarda ningún patrón significativo, por lo que la variable número de ordenadores no contribuye a la dependencia de la variable CO₂. Tanto los datos como la distribución de las categorías de la variable cualitativa "new_ordenadores" se reparten por todos los cuadrantes. Fijándonos en el vector, este presenta poca longitud, indicativa de que su representación no contiene la mayor parte de la variabilidad.



Gráfica 86: Gráfico Biplot representando agrupamiento por hora de clase.

En la Gráfica 86 las elipses, correspondientes a la hora de clase de la jornada escolar, se superponen unas con otras. Solo podemos ver las elipses de las 4ª y 5ª hora de clase algo elevadas sobre el resto. Probablemente en otra dimensión el efecto detectado para estas horas de clase en los análisis previos se vea con más claridad. En esta dimensión la perspectiva de las elipses dificulta ver bien este efecto ya constatado. Destaca la elipse de recreo que aparece con una gran extensión y alejada del origen de coordenadas.



Gráfica 87: Gráfico Biplot representando agrupamiento por número de personas.

Respecto al volumen de la estancia (Gráfica 87), el agrupamiento de datos de la categoría de 400 a 500 m³ está muy próxima al extremo del vector Volumen de la estancia y muy alejada del vector CO₂. Conforme disminuye el volumen de la estancia, más nos aproximamos al vector CO₂, quedando diferenciado como un grupo independiente en rojo aquellas que presentan volúmenes por debajo de los 100 m³, las cuales ya vimos que están ocupadas por un bajo número de personas. Esta gráfica ilustra nuevamente la correlación negativa entre volumen de la estancia y concentración de CO₂.