

# Visualización de datos multivariados



## DATA VISUALIZATION



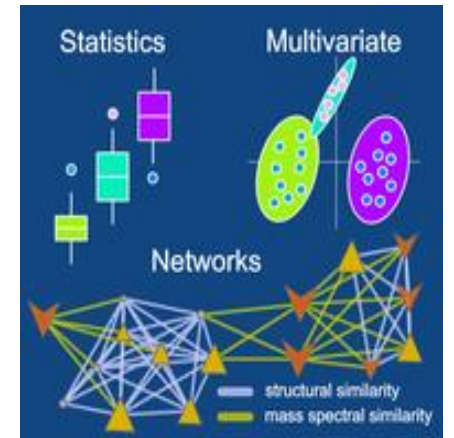
Oscar Centeno Mora

# Introducción

- ¿Por qué estudiamos la visualización de datos multivariados?
- Nos referimos a la frase de Chambers, Cleveland, Kleiner, y Tukey (1983):

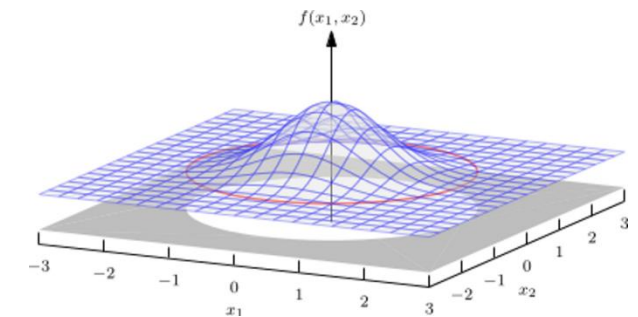
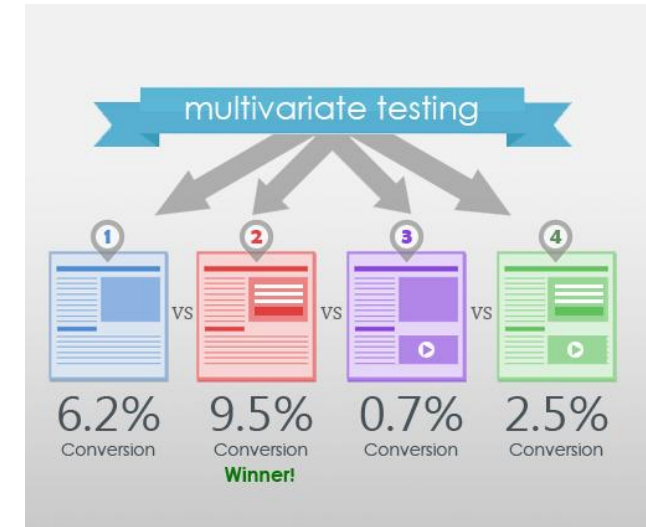
“there is no statistical tool that is as powerful as a well-chosen graph”

- En los análisis multivariados, los gráficos tienden a poseer mayores ventajas que los cuadros.
- Se crea un mayor interés y atracción en el análisis de los datos.



# Introducción

- ¿Qué entendemos por gráficos multivariantes?:
- Los gráficos se conciben como cantidades o medidas que utilizan puntos, líneas, un sistema de coordenadas, números, símbolos, palabras, sombreados, colores, etc., etc., etc.
- ¿Cuáles tipos de gráficos para una variable conocen?
- Y, ¿cuáles tipos de gráficos multivariados conocen?



# Introducción

- Algunas ventajas de los gráficos son:
- En comparación con otros tipos de representaciones, los gráficos bien diseñados crean interés y apelan a la atención del lector.
- Las relaciones visuales presentadas por gráficos son más fáciles de entender y recordar.
- El uso de gráficos ahorra tiempo, ya que el significado esencial de los datos estadísticos se pueden visualizar en un instante.
- En comparación de los cuadros, los gráficos aportan una visión más completa y general del problema en causa.



# Introducción

- Un gráfico multivariado persigue 4 objetivos esenciales:

1. Proporciona una visión general.

2. Relata una historia.

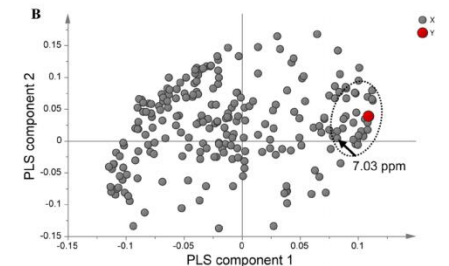
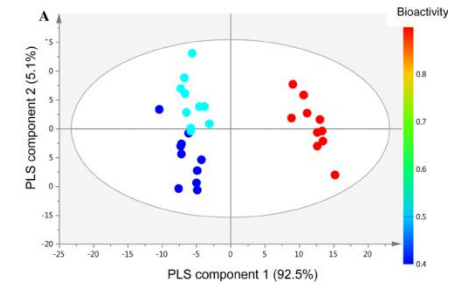
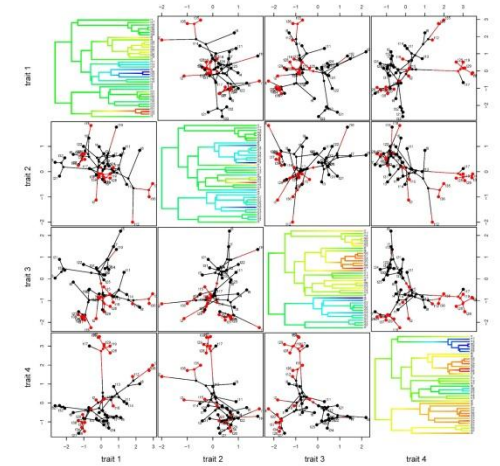
3. Sugiere una hipótesis.

4. Critica o aprueba un modelo.



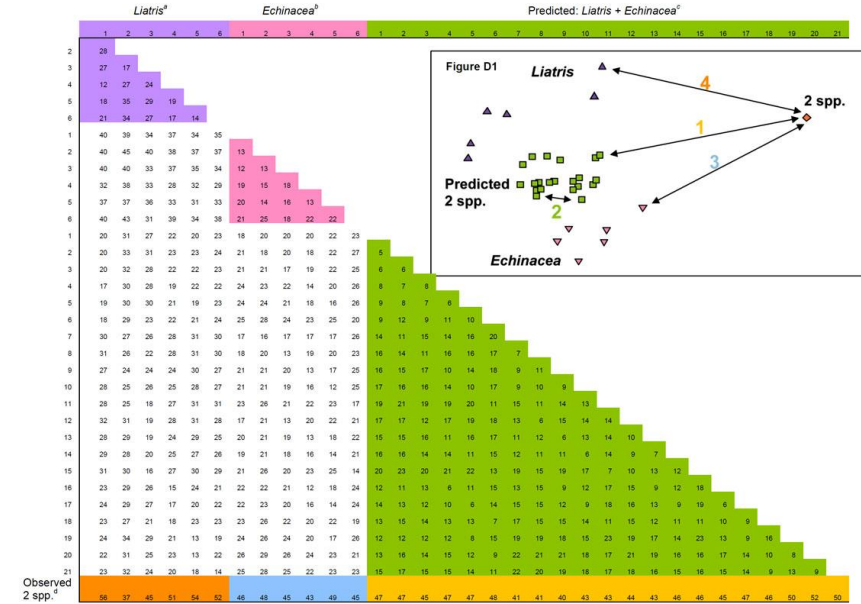
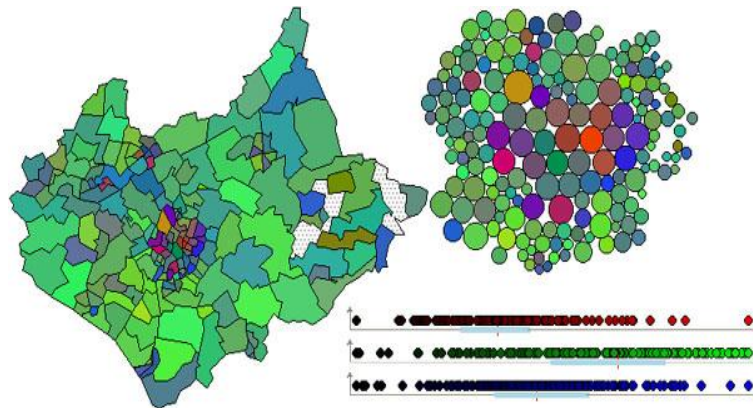
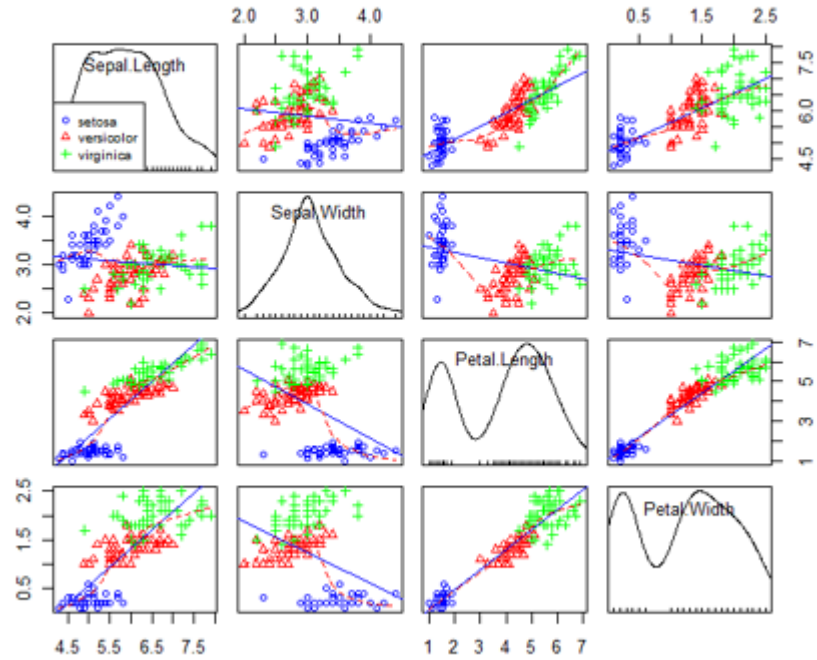
# Introducción

- El presente capítulo se enfoca en estudiar los gráficos multivariados más importantes.
- Entre estos se encuentran: matrices de gráficos, correlogramas, gráficos tridimensionales, contornos, densidades, mosaicos, etc.
- Nótese que existen gráficos que muestran más allá de su esencia base, por ejemplo, se puede visualizar la variabilidad para gráficos donde se busca estudiar la posición.
- Antes de empezar cualquier tipo de análisis estadístico, debemos comenzar por estudiar, caracterizar y analizar los datos, y que mejor forma que con gráficos.

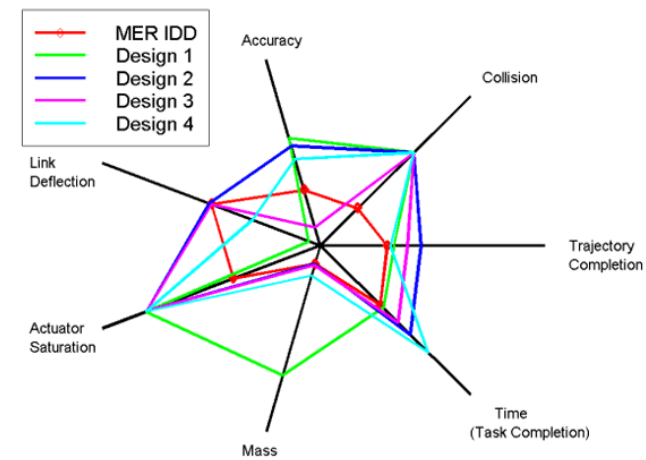




## Iris Data

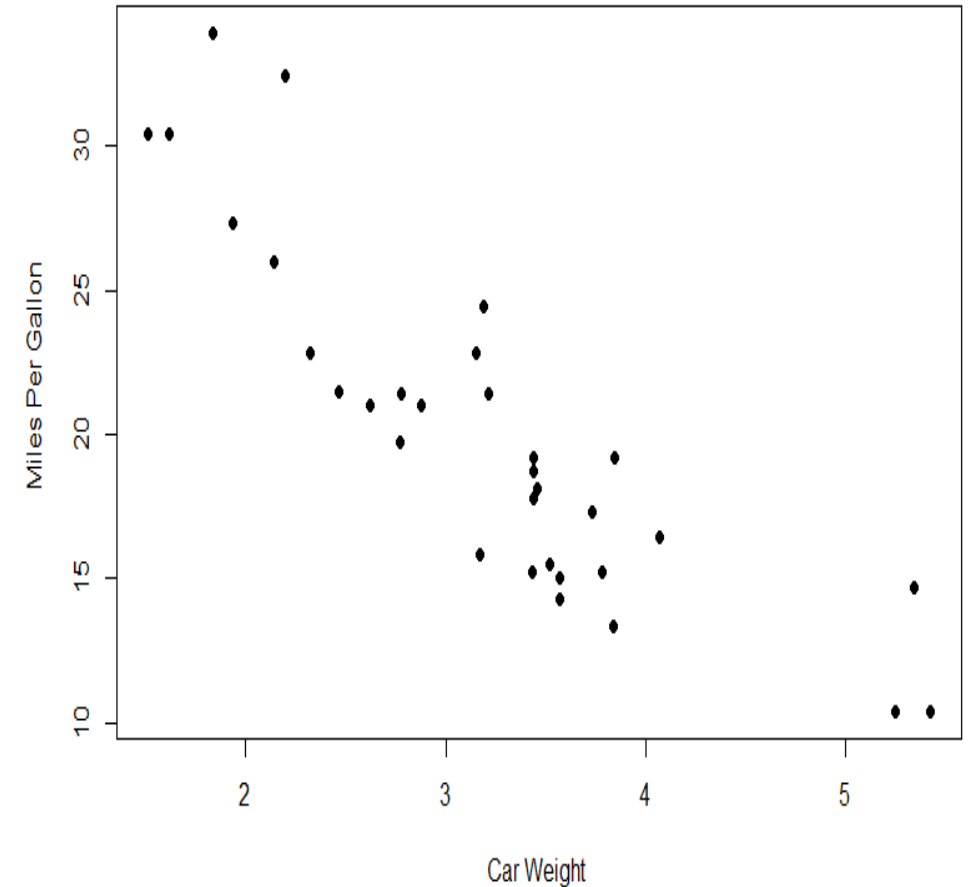


Star Plot of MER IDD and Automated Designs



# Gráficos de puntos

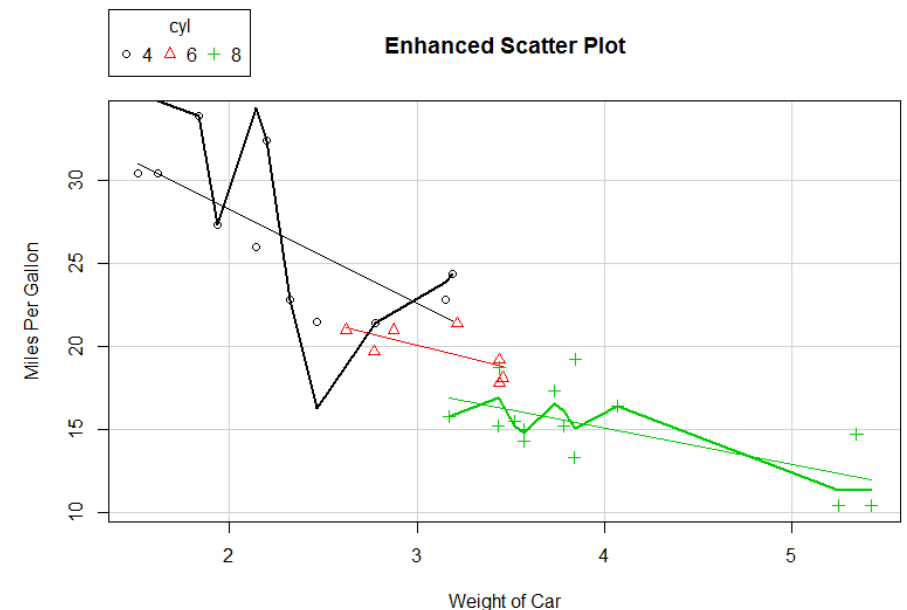
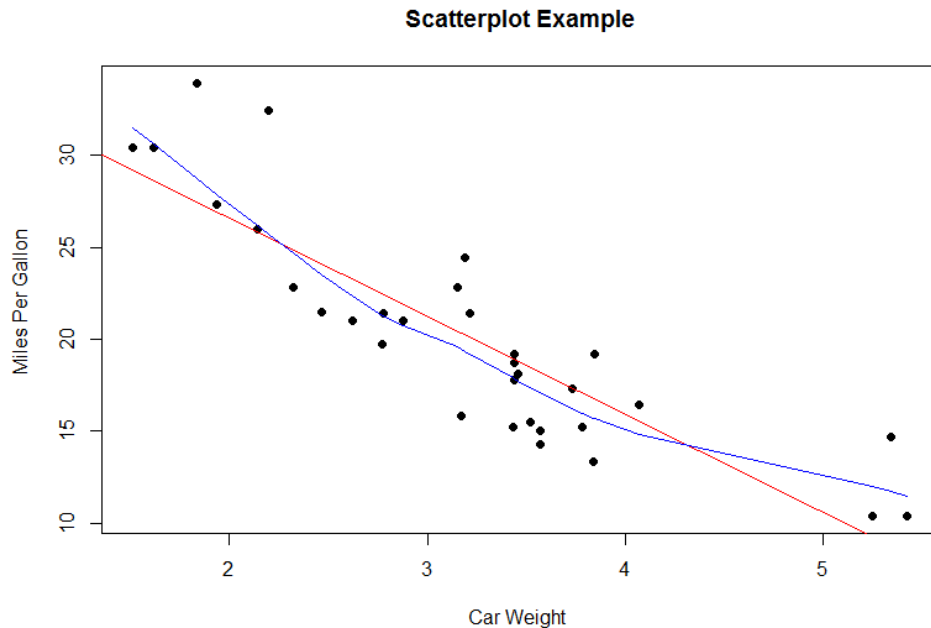
- El objetivo de los gráficos de puntos bivariados es encontrar relaciones en las variables.
- ¿Para qué tipo de variables se utiliza?
- ¿Cómo interpretaríamos el siguiente gráfico?
- ¿Qué se puede decir del peso y las millas por galon?
- ¿Se podría mejorar el entorno gráfico?





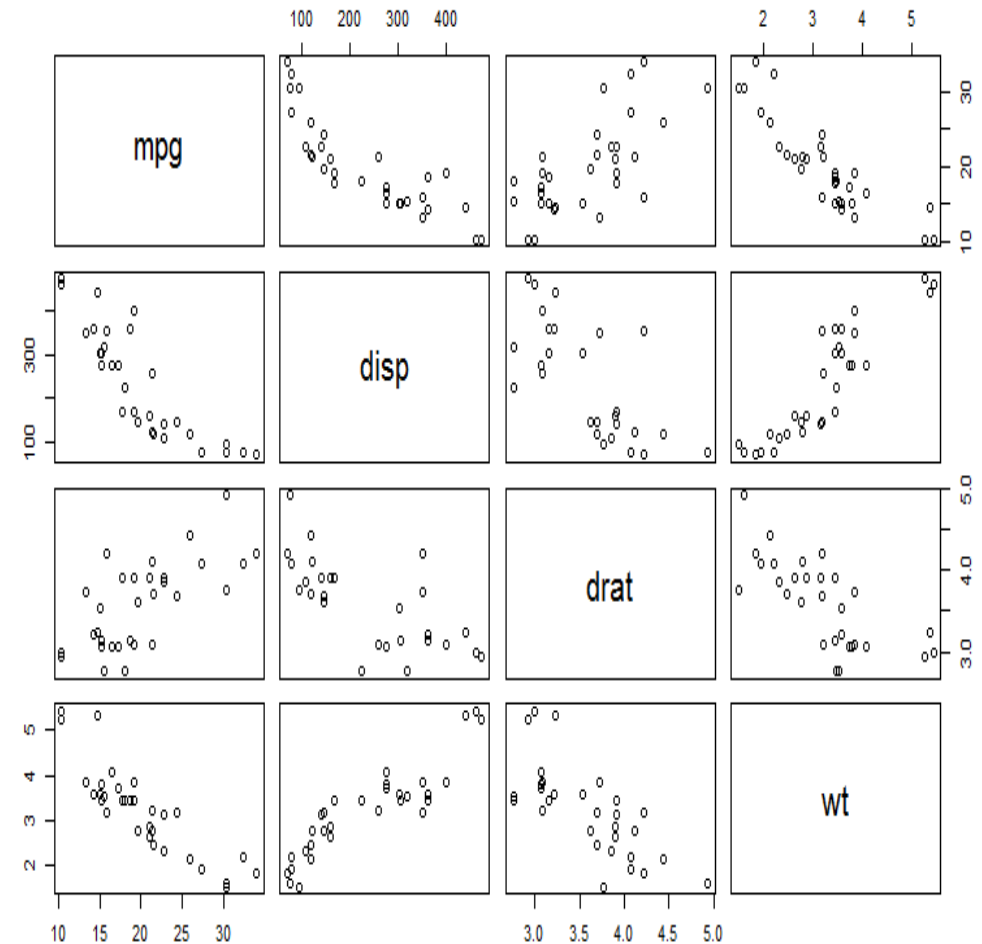
# Gráficos de puntos

- Los gráficos de puntos bivariados pueden ser más explícitos e informativos en lo que a la tendencia se refiere.
- También podemos separar las tendencias según las categorías implícitas de los datos.



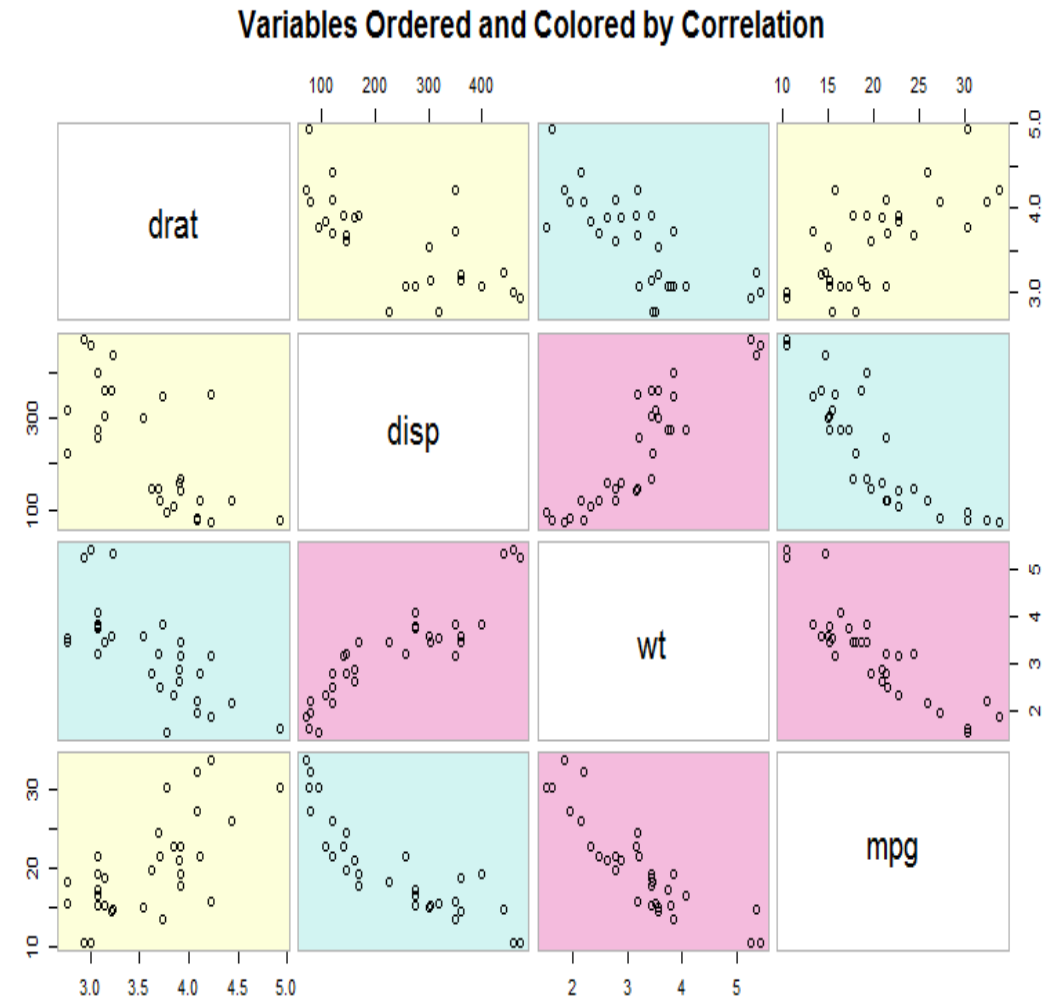
# Matrices de puntos

- Se utilizan también en el estudio de las correlaciones, en variables cuantitativas.
- El estudio conjunto permite describir mejor el comportamiento conjunto de las variables.
- También, permite aproximar el tipo de relación y la intensidad de la misma.
- Este análisis es muy común en los análisis de regresión.
- ¿Qué podemos decir del presente ejemplo?



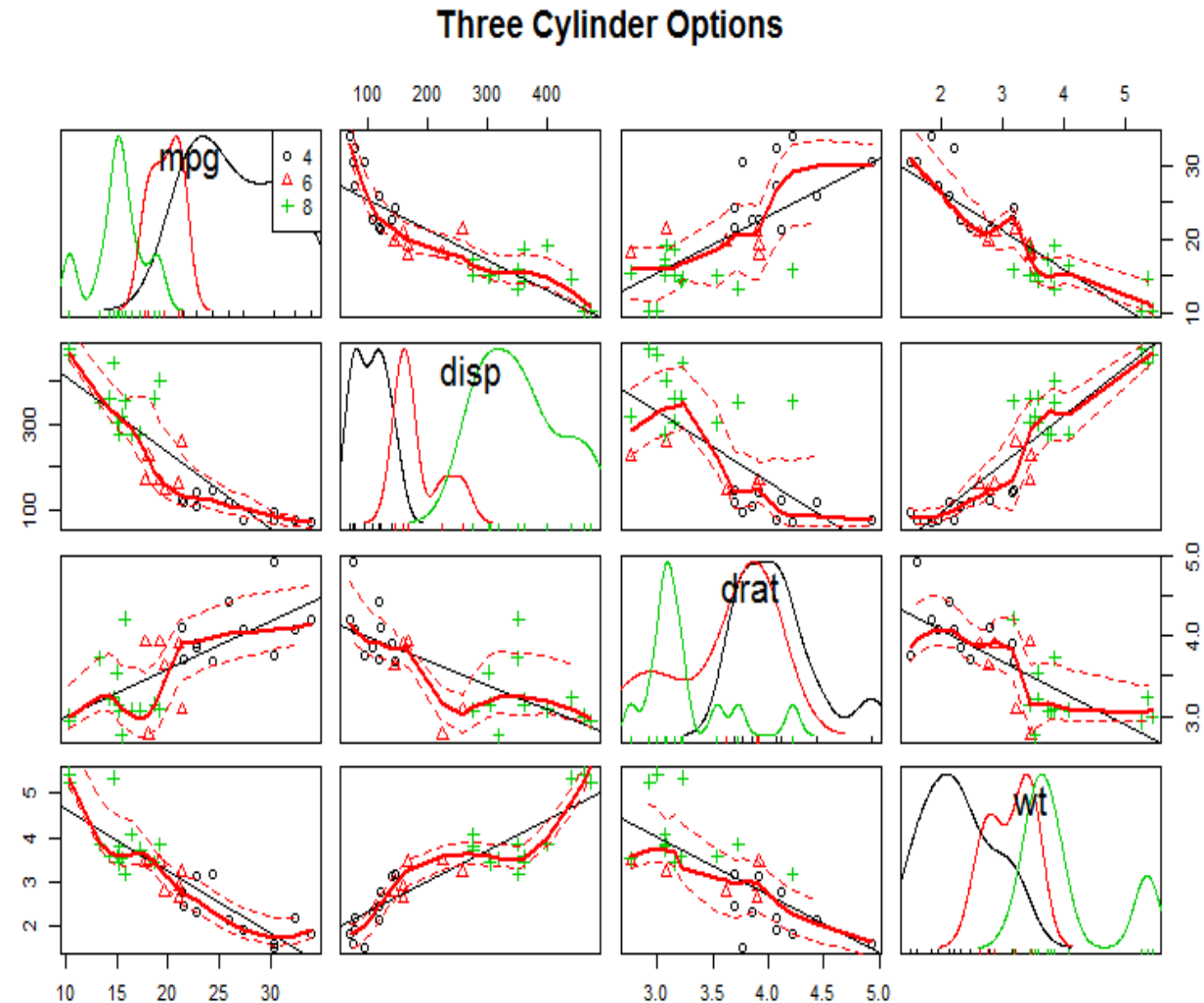
# Matrices de puntos

- Para mejorar la visualización respecto al tipo de relación para las variables, según, tipo de relación, intensidad, etc., una opción son los colores.
- Colores intensos pueden reflejar la fuerza o magnitud de la correlación, y entre menor la relación, el color se irá degradando.
- ¿Qué podemos decir del presente caso?



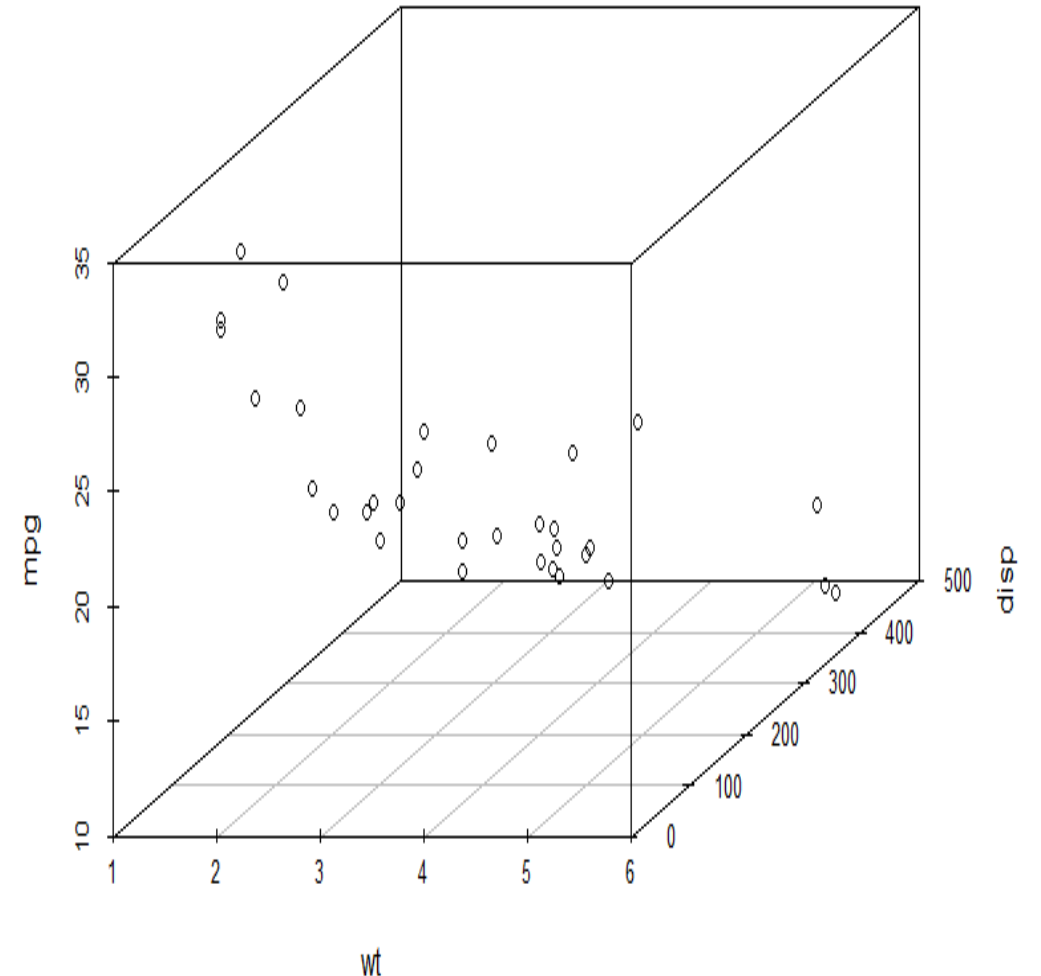
# Matrices de puntos

- Los gráficos de puntos pueden ser aún más informativos, y no se limitan a una simple relación entre los puntos.
- Por ejemplo, se puede mostrar la distribución de los datos para cierta variable, se pueden meter categorías, tendencias para cada una de las categorías, etc.
- ¿Qué se podría decir del presente gráfico?



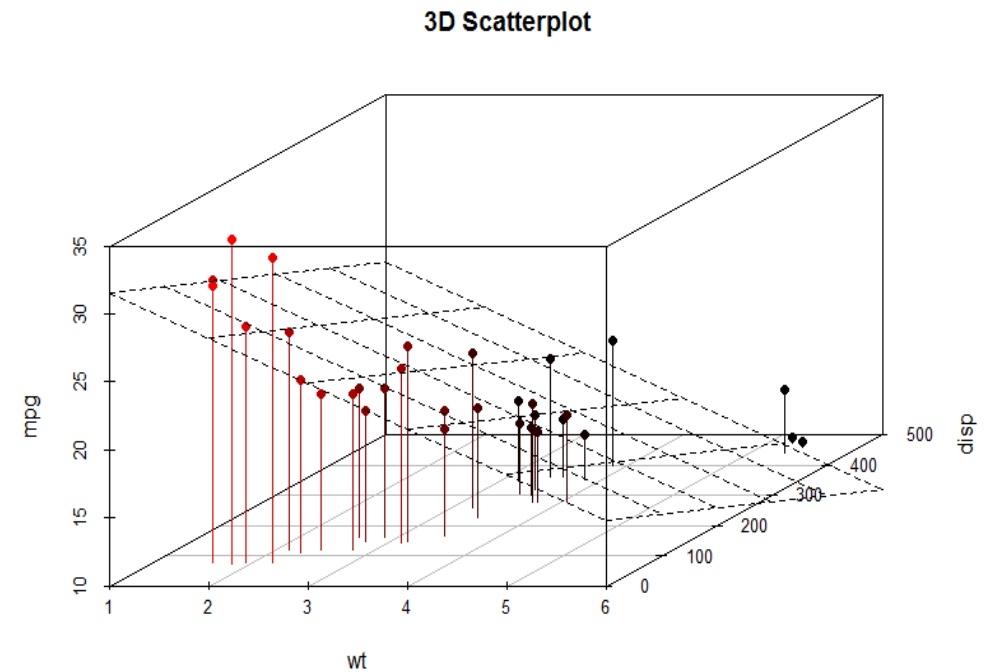
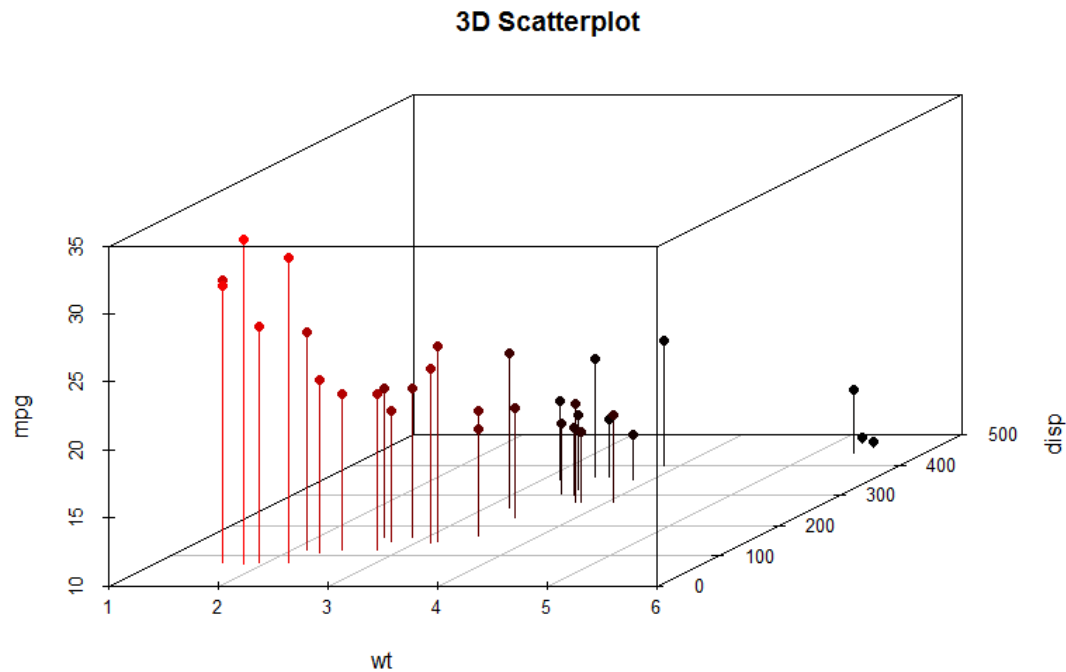
# Gráficos de puntos tridimensionales

- Otra forma de conocer la relación entre otras variables, son puntos puestos en tres dimensiones.
- Conocer la interacción conjunta de tres variables puede ser un insumo importante para el análisis.
- Sin embargo, ir más allá de 2 dimensiones es una labor a veces difícil de comprender.
- ¿Qué podemos decir de la presente representación?



# Gráficos de puntos tridimensionales

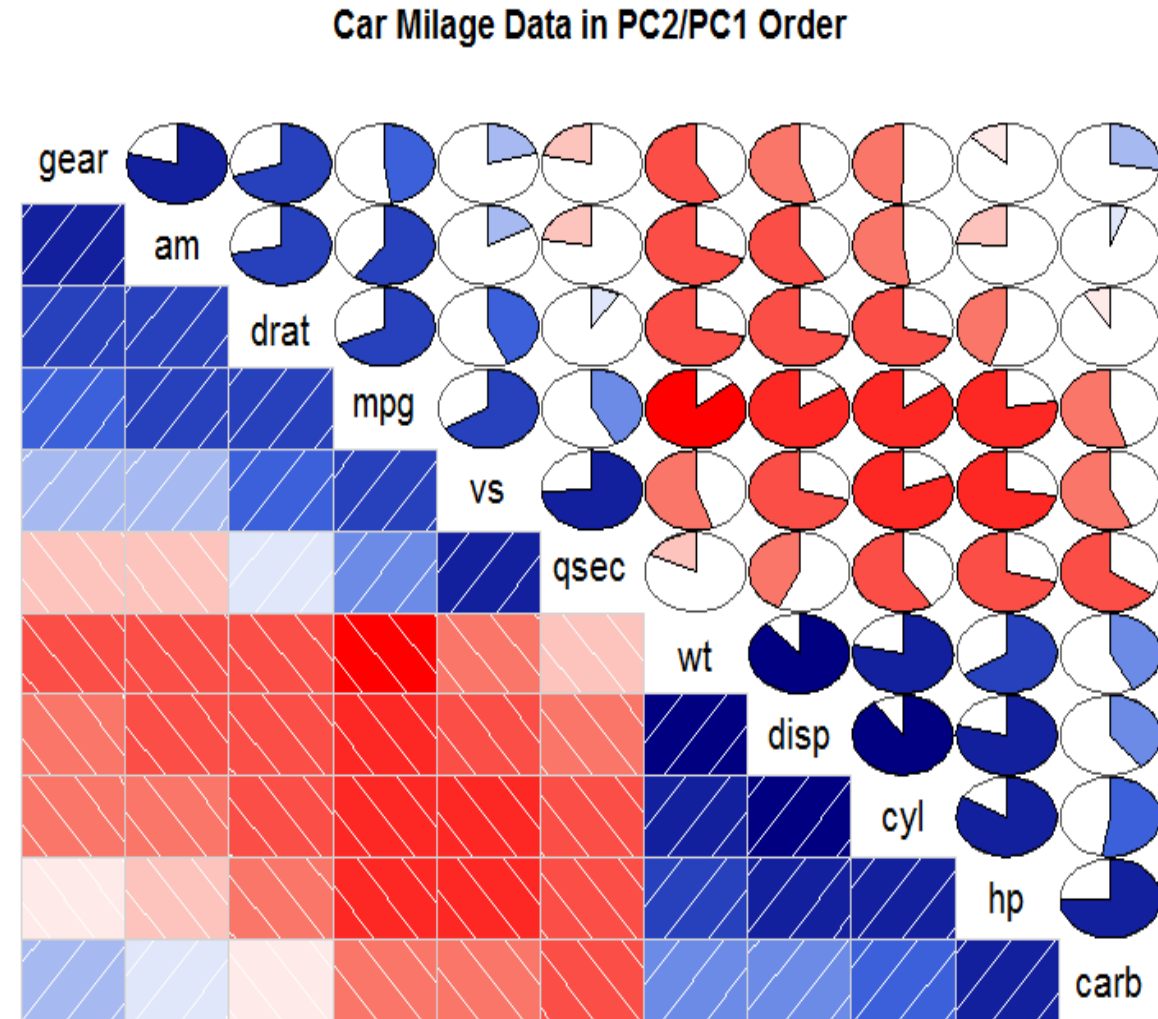
- Una forma de guiar el análisis es utilizando bastones o líneas del punto a la superficie, o híper planos para conocer la dirección o la relación entre las variables.





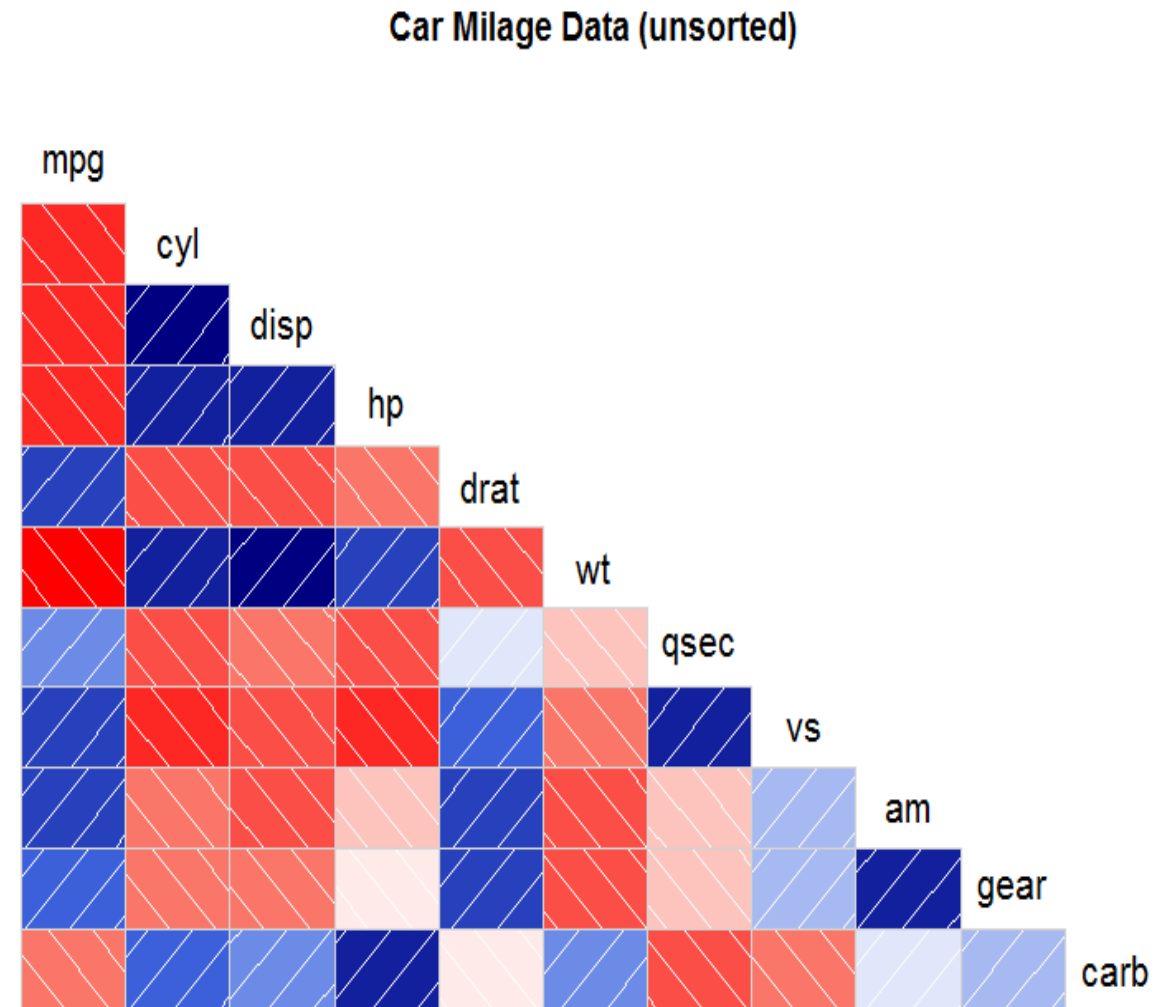
# Correlograma y los colores

- Para una mayor cantidad de variables, una forma atractiva es mediante un correlograma.
- Notar que los correlogramas pueden existir para variables de corte transversal como longitudinal.
- En este tipo de correlograma se puede apreciar tanto el tipo de relación, la magnitud de la relación, así como diversos indicadores que mejoran el análisis de los datos.
- ¿Qué se podría decir del siguiente análisis?



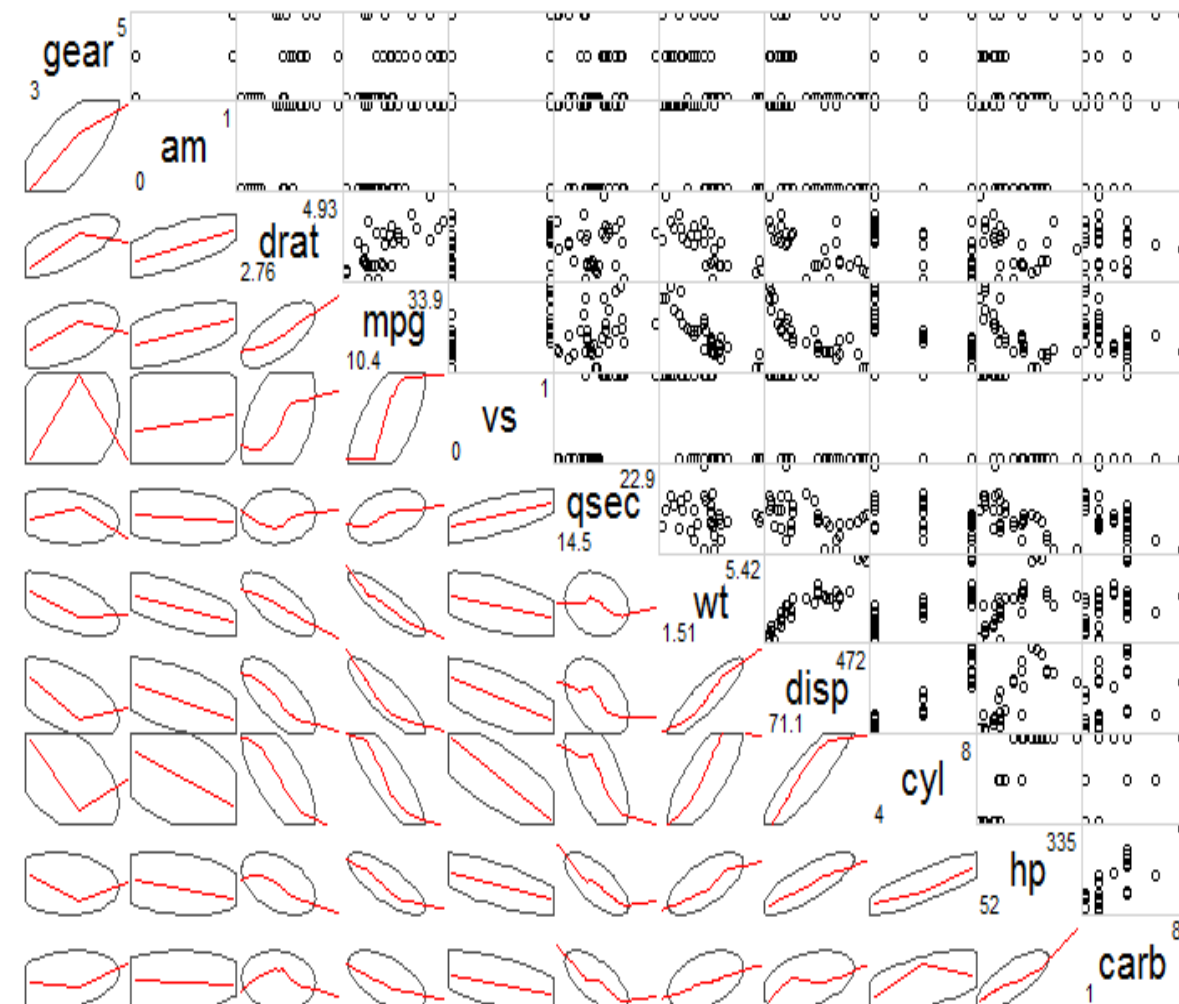
# Correlograma y los colores

- El correlograma anterior, se ordenó según magnitudes.
- Se puede ordenar las variables para tratar de mostrar otro tipo de caracterización.



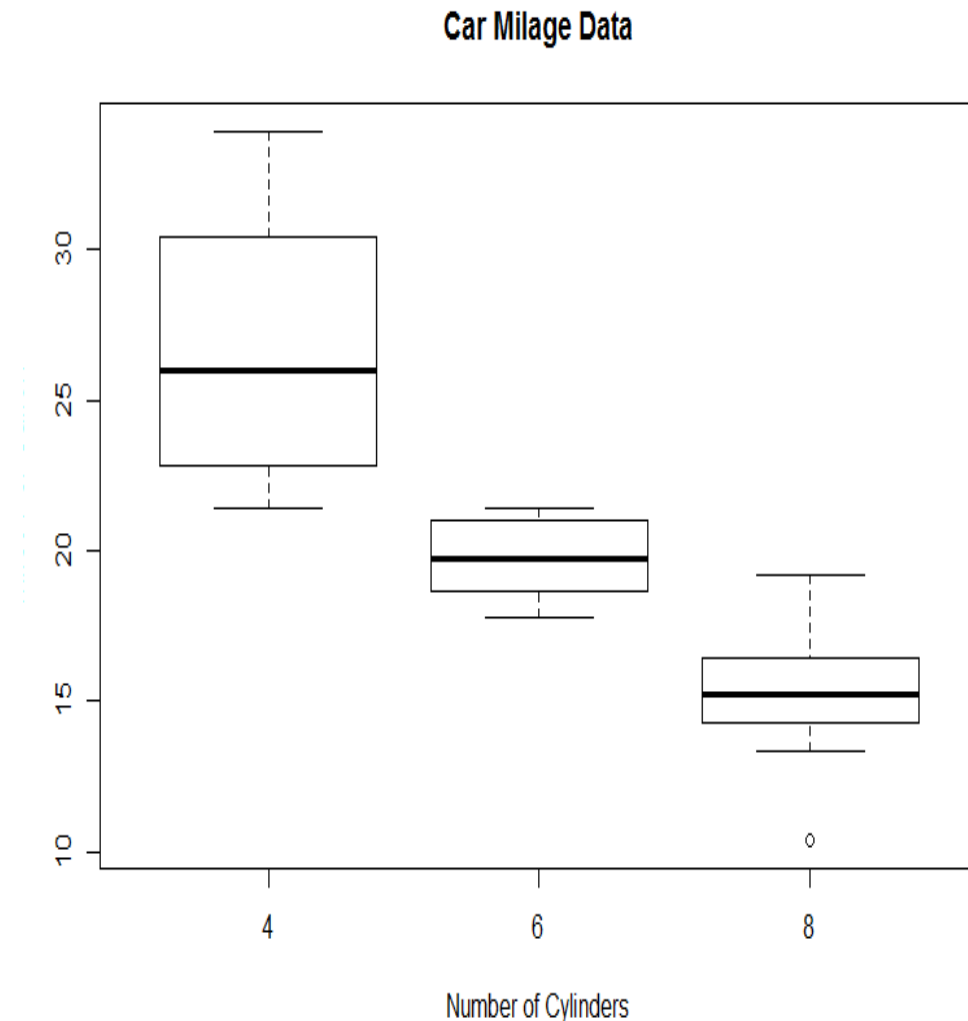
# Correlograma

- También se puede definir de la forma típica, puntos y líneas.
- Esta forma es muy útil para un analista, donde prefiere saber con mayor exactitud el tipo de relación y la forma.
- El anterior puede servir más para personas en puestos de decisión, donde no les interese profundizar en los análisis.



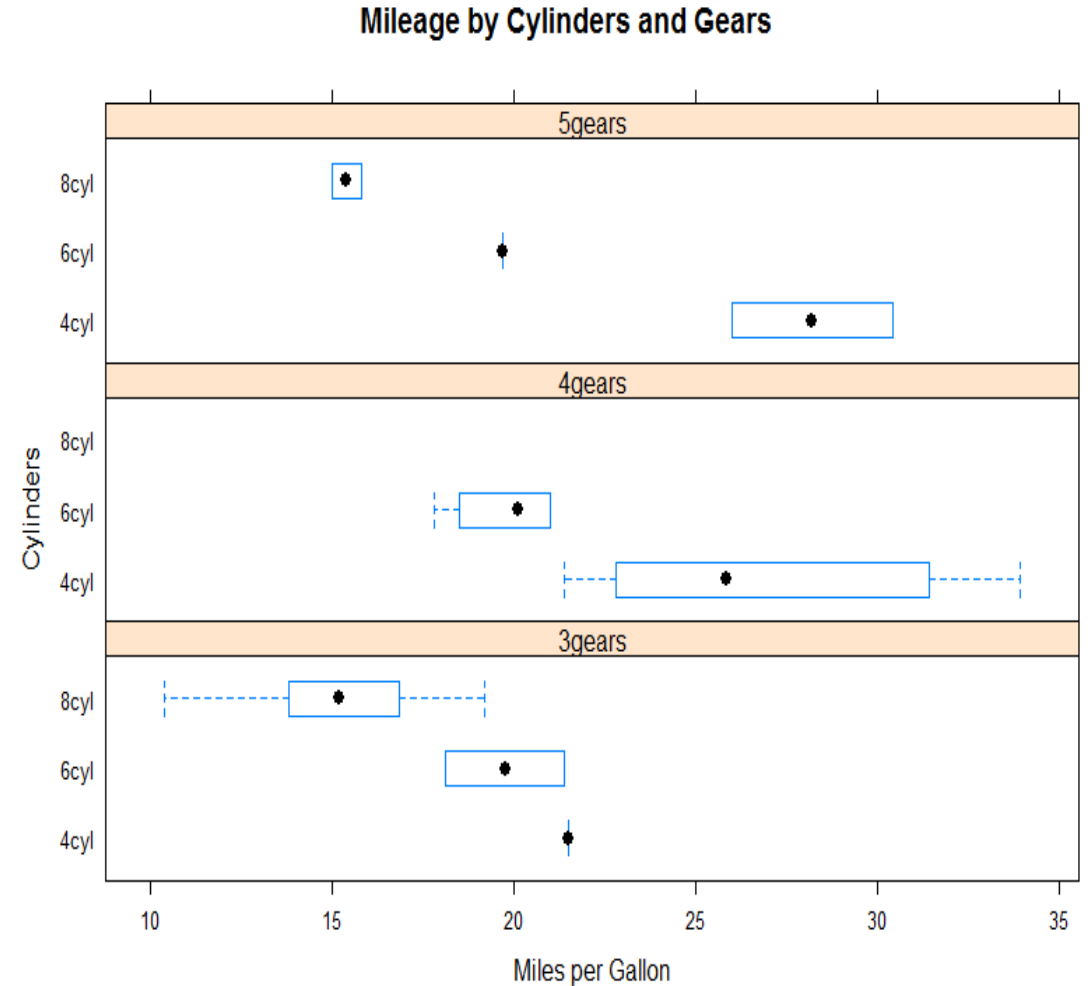
# Gráficos de cajas

- El gráfico de cajas o boxplot sirve para analizar tanto tendencia como dispersión en un conjunto de datos.
- Sirve para analizar variables cuantitativas.
- Técnicas como Análisis de Variancia y otros se sirven de este tipo de gráficos.
- En la comparación entre categorías, es un gráfico muy recomendado.



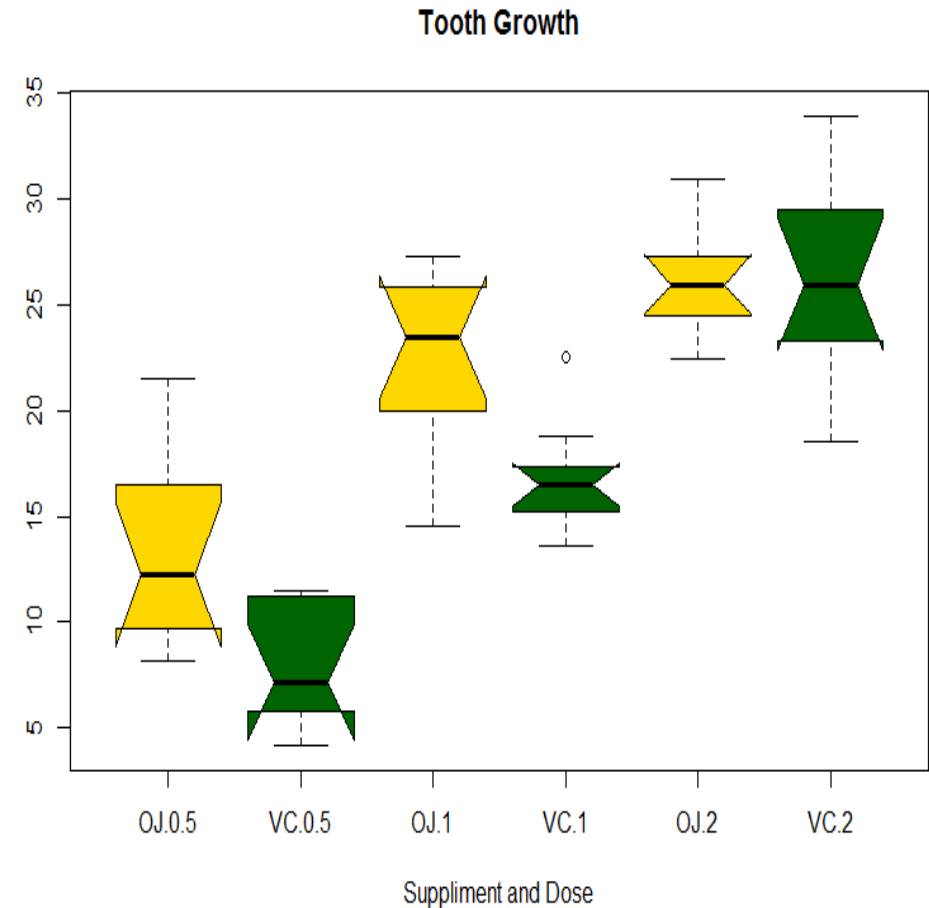
# Gráfico de cajas

- Los gráficos de cajas se pueden cruzar para más de dos variables.
- Pueden ser en dirección vertical u horizontal.
- ¿Qué podemos decir de marcha de 4 “4gears”?
- ¿Qué podemos decir de todo el gráfico?



# Gráficos de cajas

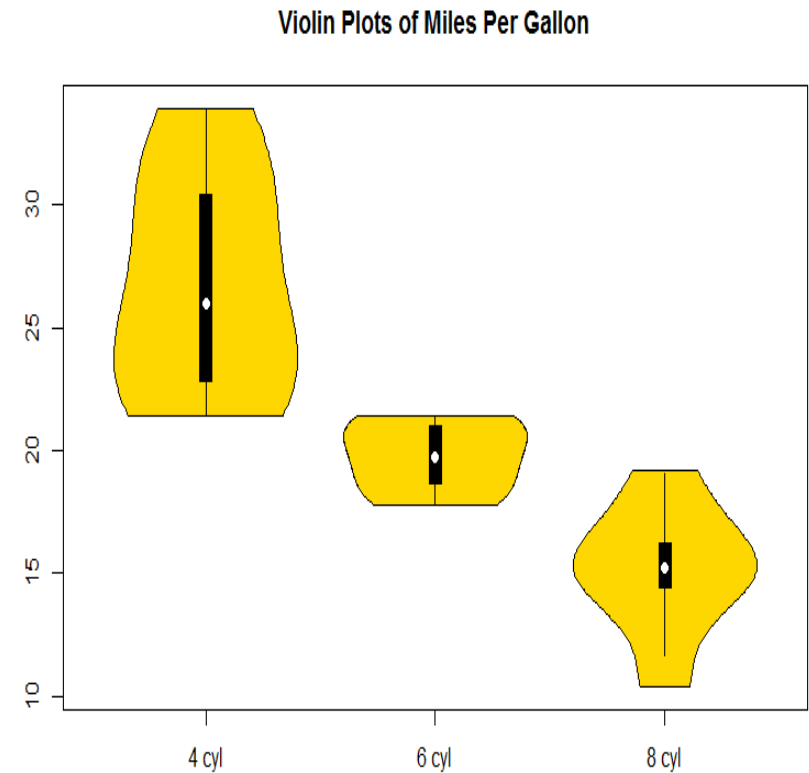
- El gráfico de cajas puede presentarse de diversas formas, esto con el fin de hacer más llamativo la figura.
- La funcionalidad es la misma, solo que acá se puede apreciar mejor las asimetrías, los valores extremos, y la comparación entre las categorías.
- ¿Qué podemos decir el presente gráfico?





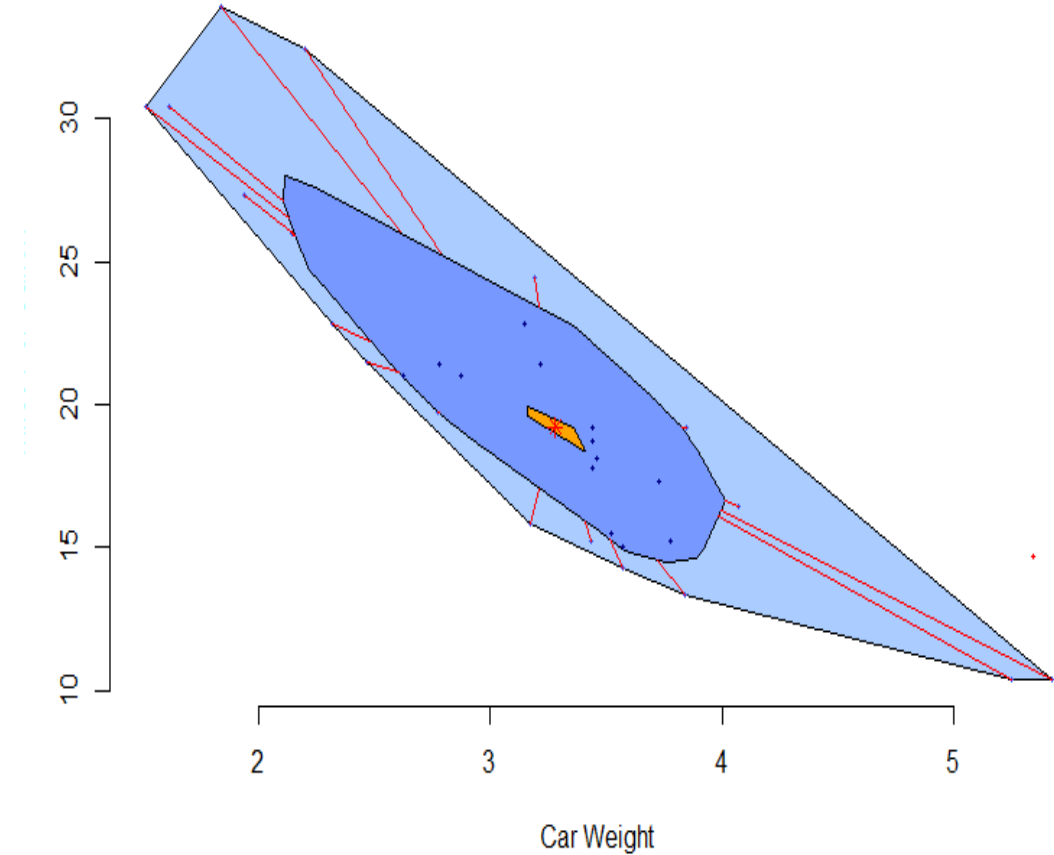
# Gráficos de cajas

- Otro tipo de gráfico de cajas son los violines.
- Este tipo de cajas busca presentar de la mejor forma la variabilidad dentro de los grupos.
- ¿Qué podemos decir el presente gráfico?



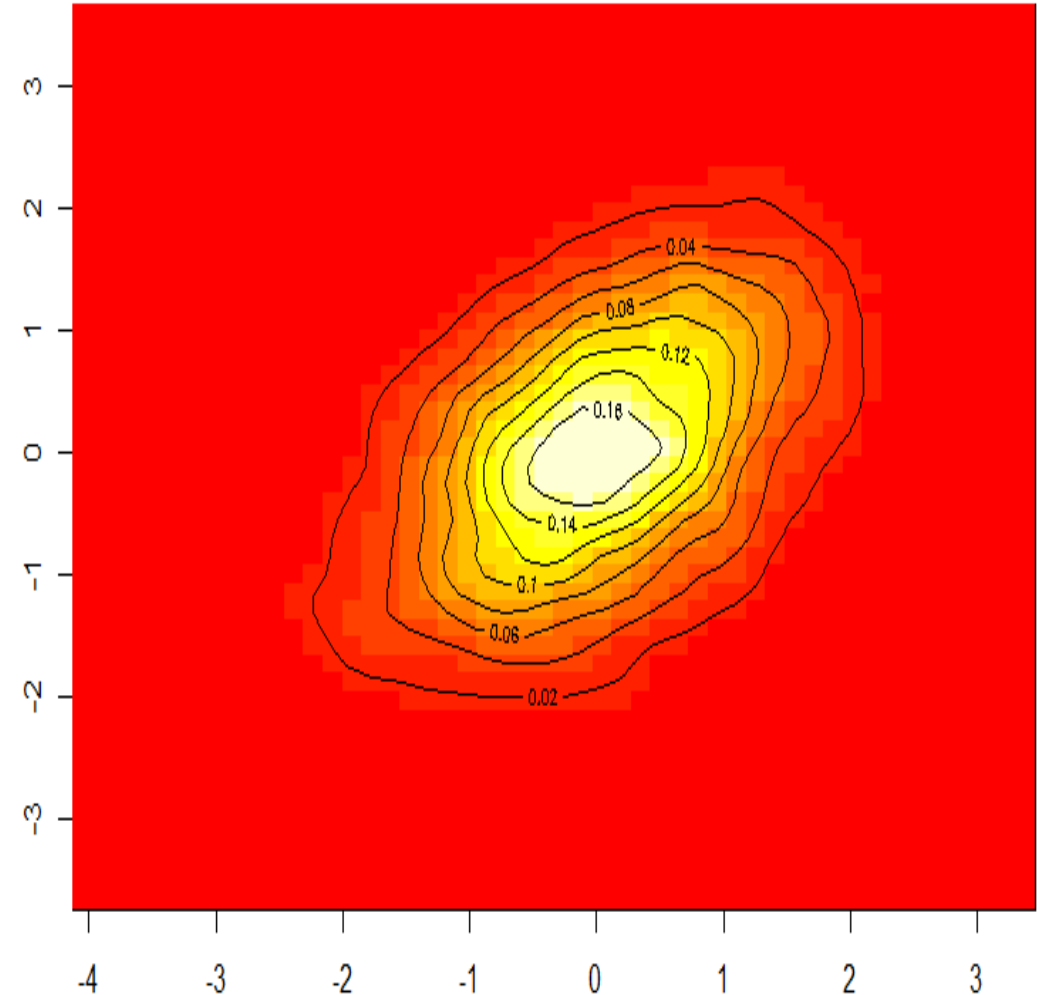
# Gráficos de cajas

- Finalmente, el último tipo de gráfico de caja es el gráfico de bolsa, o el bagplot.
- Proporciona una versión bivariada. La bolsa contiene el 50% de todos los puntos. La mediana bivariente es aproximada. La valla separa los puntos de la valla de los puntos exteriores. Se muestran los valores atípicos
- ¿Qué podemos decir el presente gráfico?



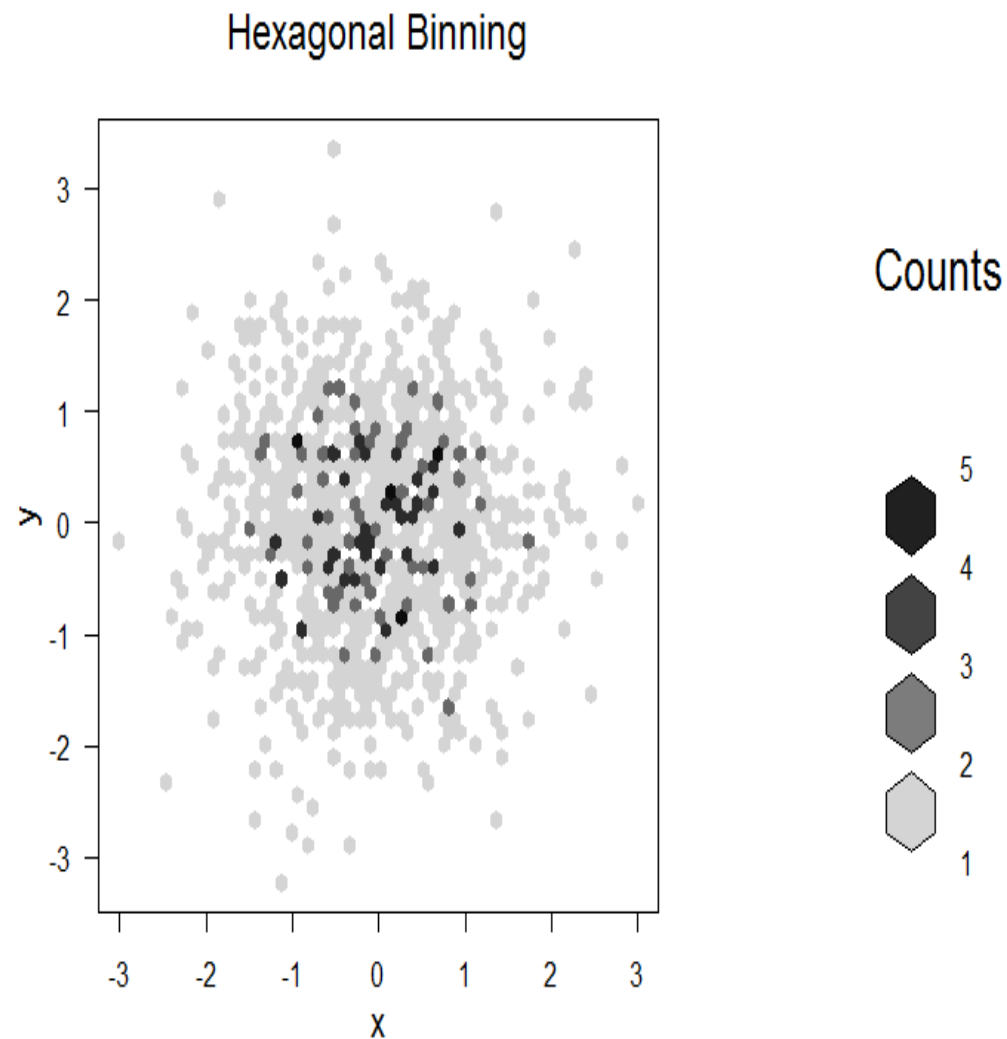
# Gráfico de contorno

- Los gráficos de contorno, y a veces conocidos como gráficos de calor , sirven para buscar óptimos en la distribución de datos.
- A veces interesa saber donde es que se encuentra el mejor punto de calentamiento, o en que nivel, presión, calibración, etc., el proceso se optimiza.
- ¿Qué podemos decir el presente gráfico?



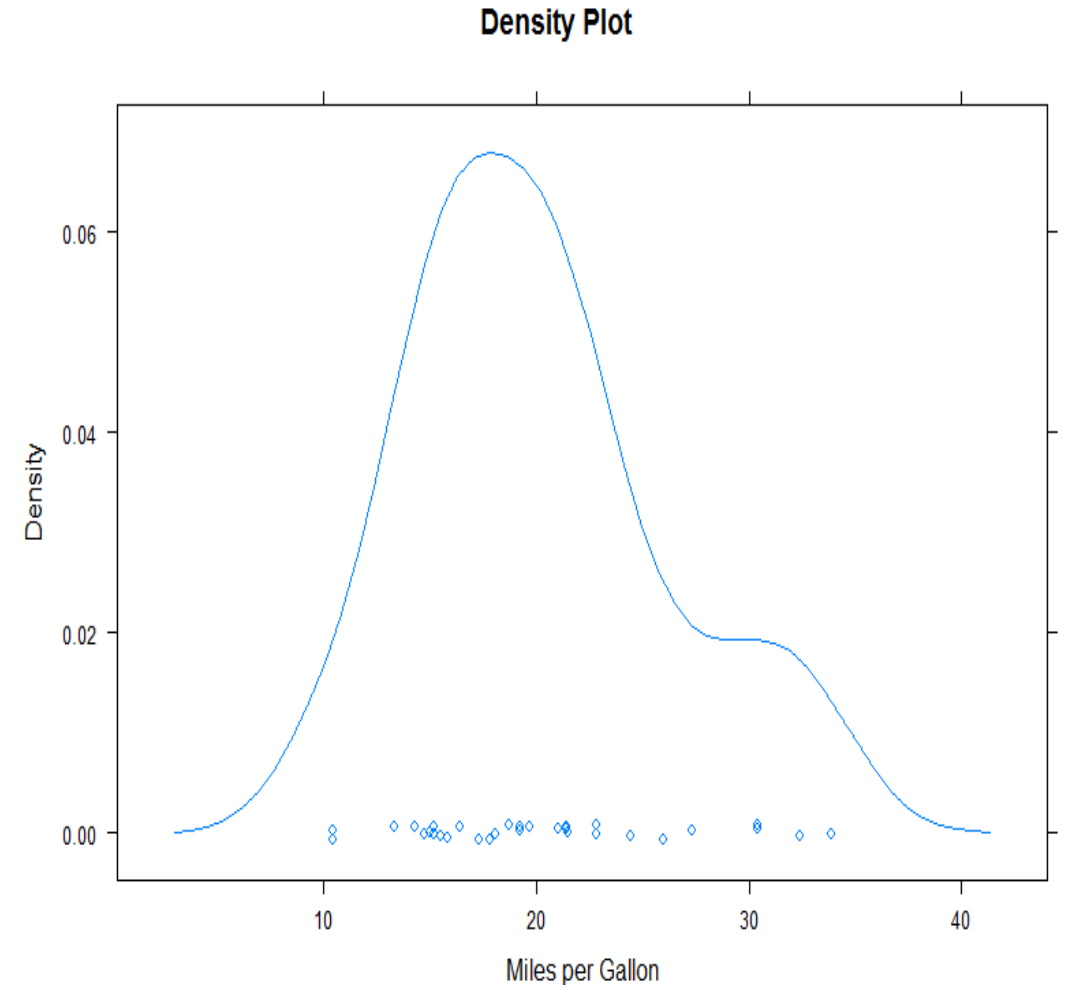
# Gráfico de contorno

- Los contornos se pueden visualizar también como conteos.
- Este formato es un poco diferente, en el sentido que no se visualizan regiones, sino que hay una pequeña categorización, y esta se ve en pequeños polígonos.
- ¿Qué podemos decir del presente gráfico?



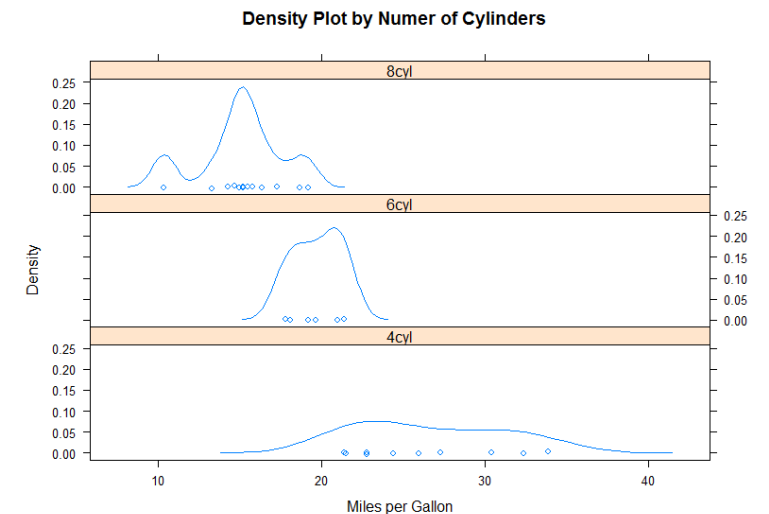
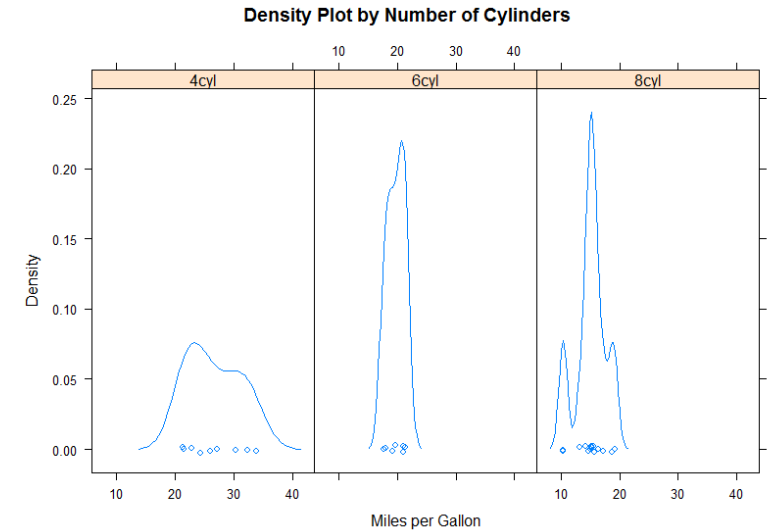
# Gráficos de densidad

- El estudio de las densidades es muy importante. Se aplica para datos cuantitativos, y sirve para saber la forma o distribución de los datos.
- Las densidades son guía esencial para saber si los datos poseen valores extremos, o si se debe hacer el supuesto de la normalidad.
- ¿Qué se podría decir de la siguiente densidad?



# Gráficos de densidad

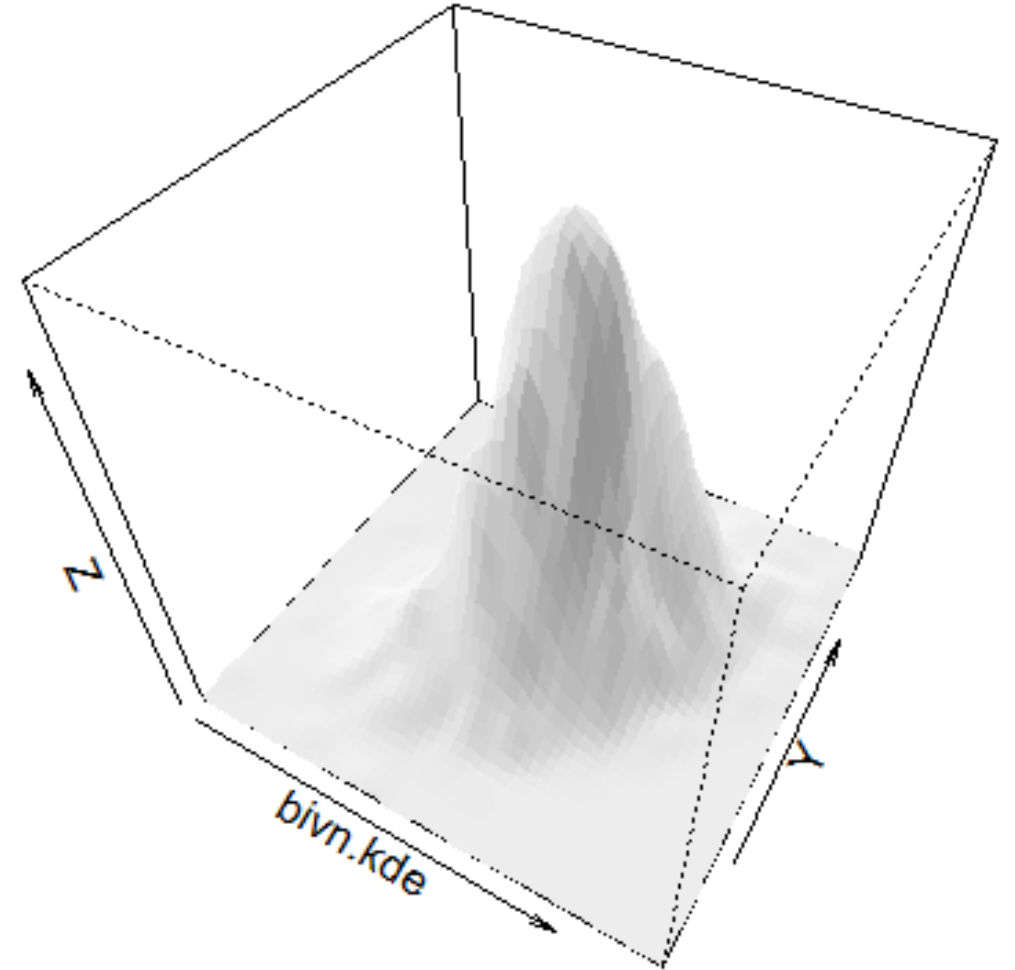
- La densidad se puede aplicar en el estudio conjunto de diversas categorías.
- Se puede apreciar tanto la posición como la forma de una distribución.
- ¿Qué se puede decir de las siguientes densidades.?





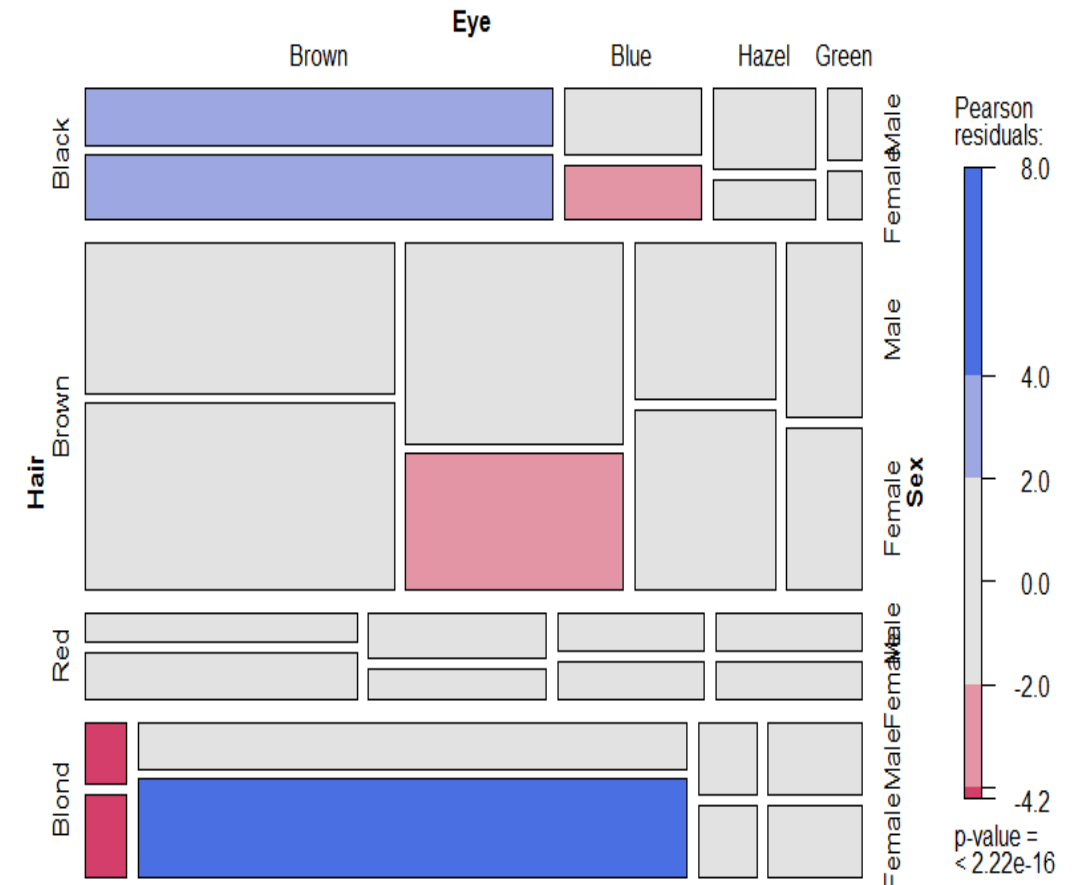
# Gráficos de densidad tridimensional

- Asimismo, se puede llevar a cabo un análisis de densidad para 3 variables.
- Aunque es difícil, se puede ver un poco las curvaturas y áreas de posición.
- ¿Qué podemos decir del siguiente gráfico?



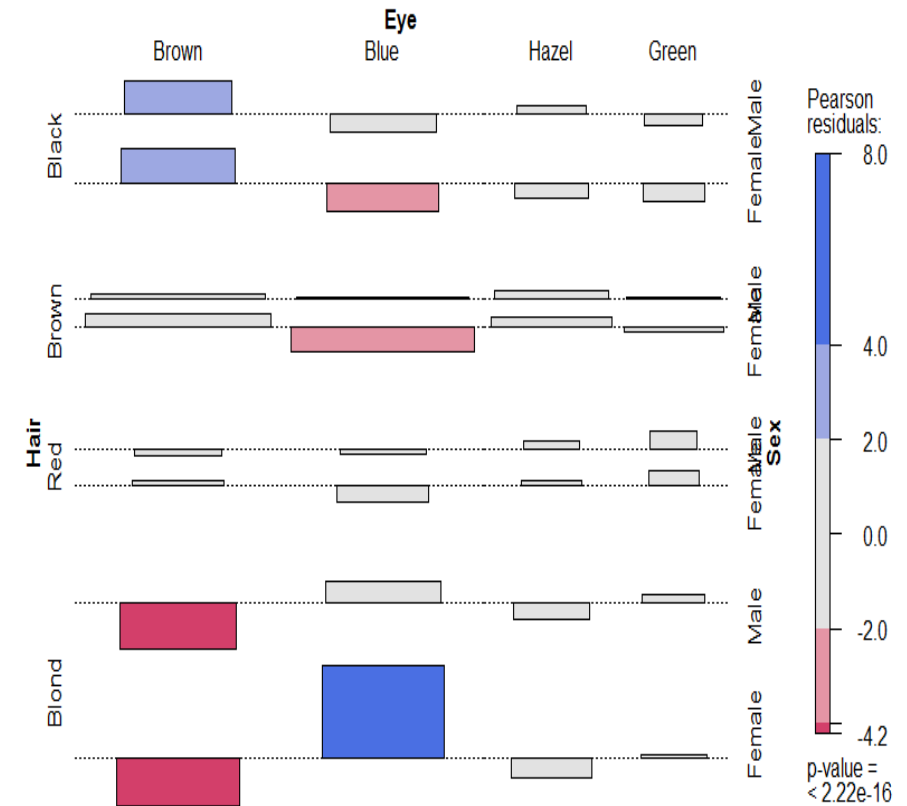
# Gráficos de mosaico

- El gráfico de mosaico sirve para visualizar datos multivariados donde se poseen variables cualitativas.
- Existen muchas formas de parametrizarlo: según las categorías conjuntas, según una categoría de referencia, etc.
- Se pueden hacer pruebas de asociación e independencia, y dibujar el gráfico según esto.
- ¿Qué podemos decir del siguiente gráfico?



# Gráficos de mosaico

- Aunque el objetivo es la descripción de los datos, se podría hacer un análisis de asociación entre las variables categóricas.
- ¿Qué podemos decir del siguiente gráfico?

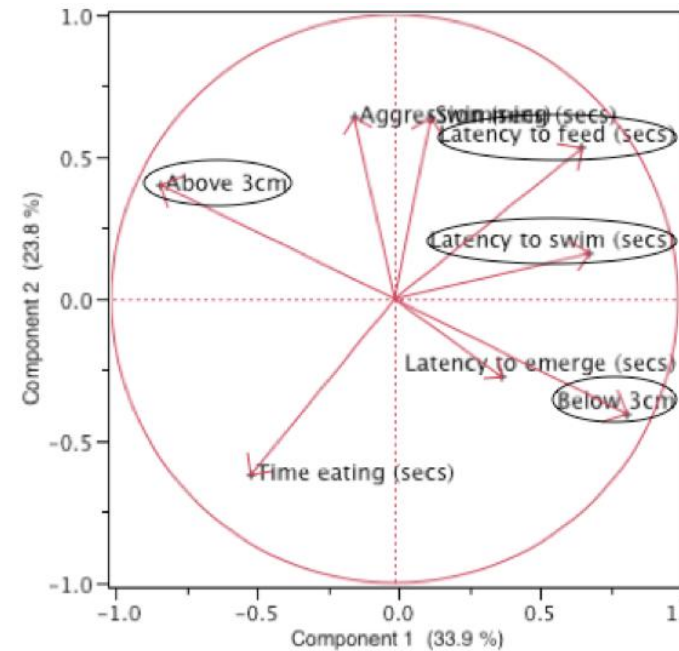
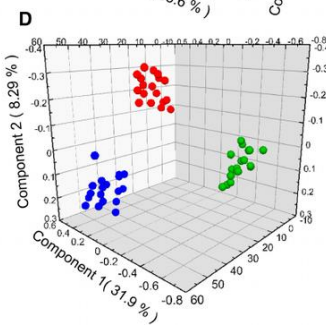
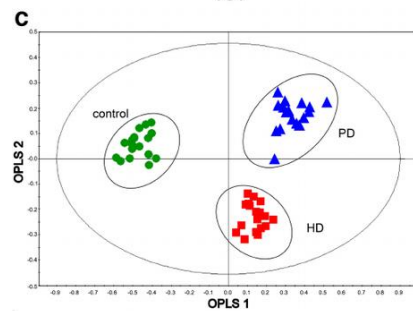
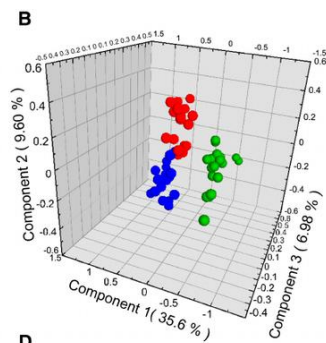
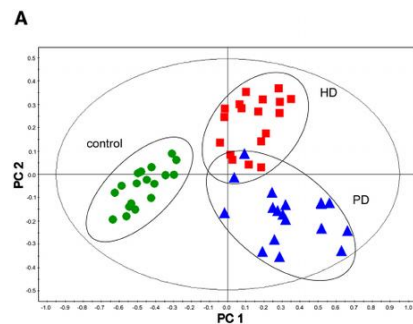


# Y, ¿los gráficos del curso?



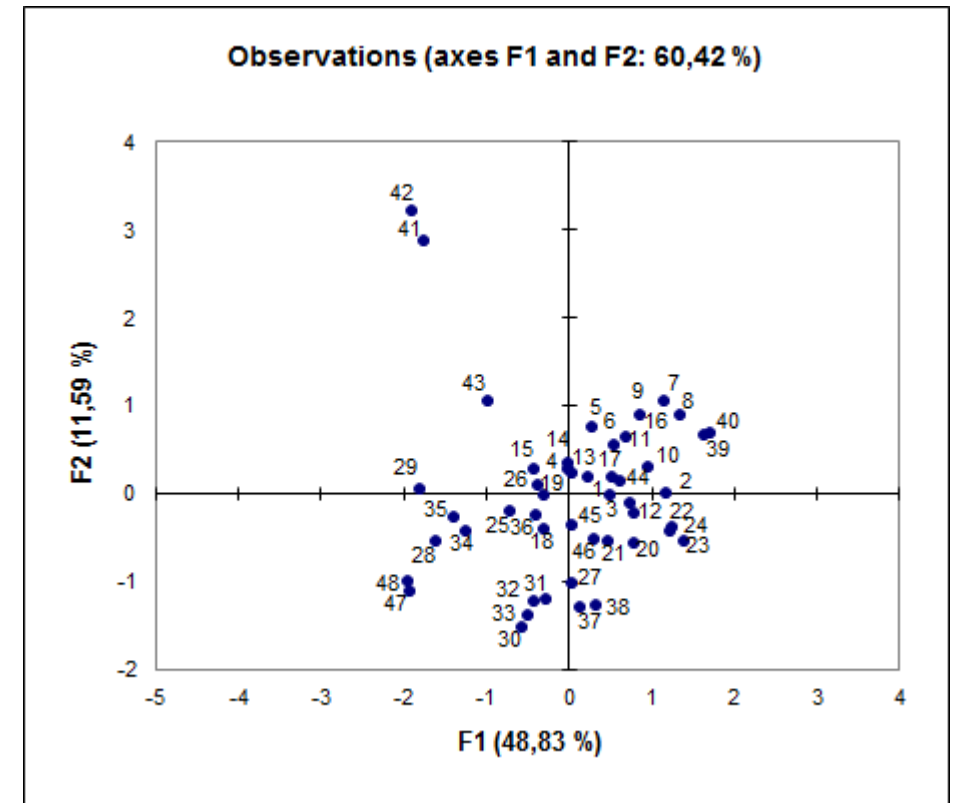
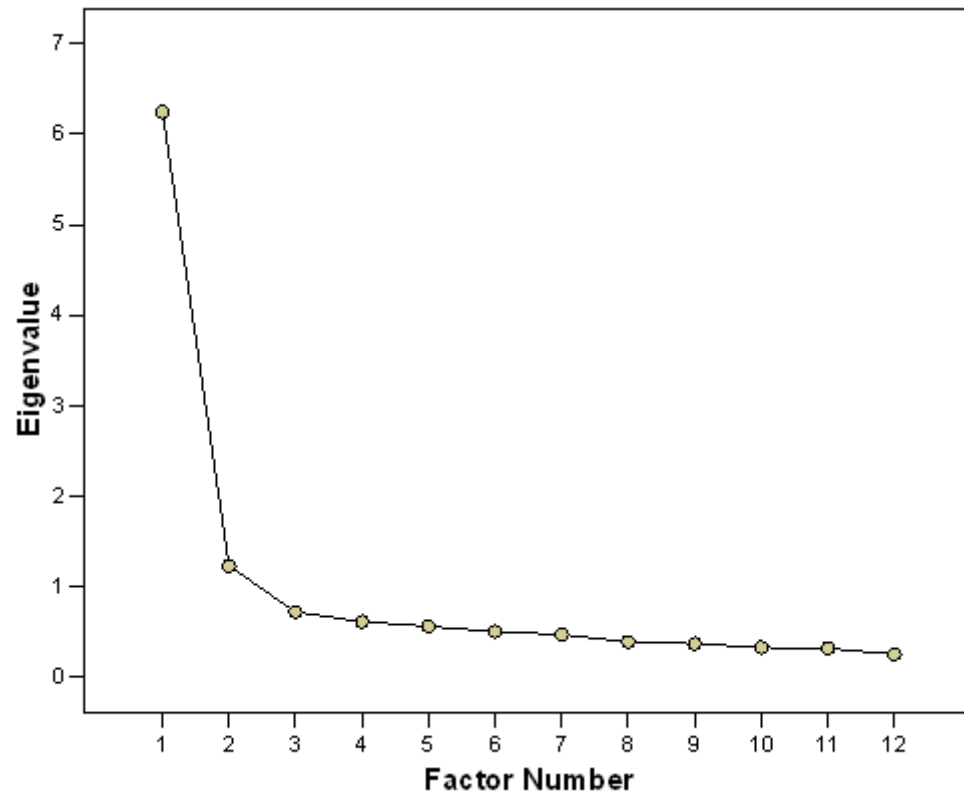
# Gráficos multivariados de las técnicas del curso

- Para el PCA.



# Gráficos multivariados de las técnicas del curso

- Para el análisis de Factores.

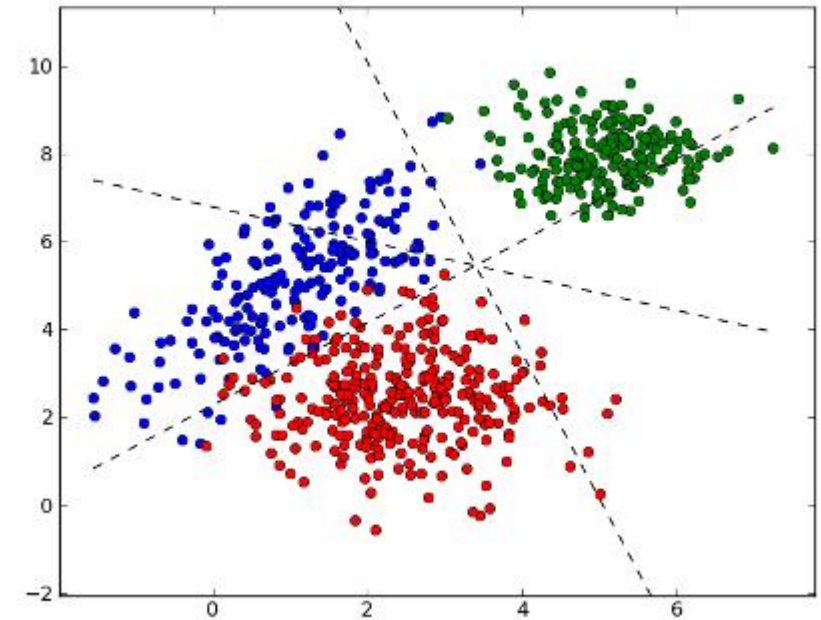
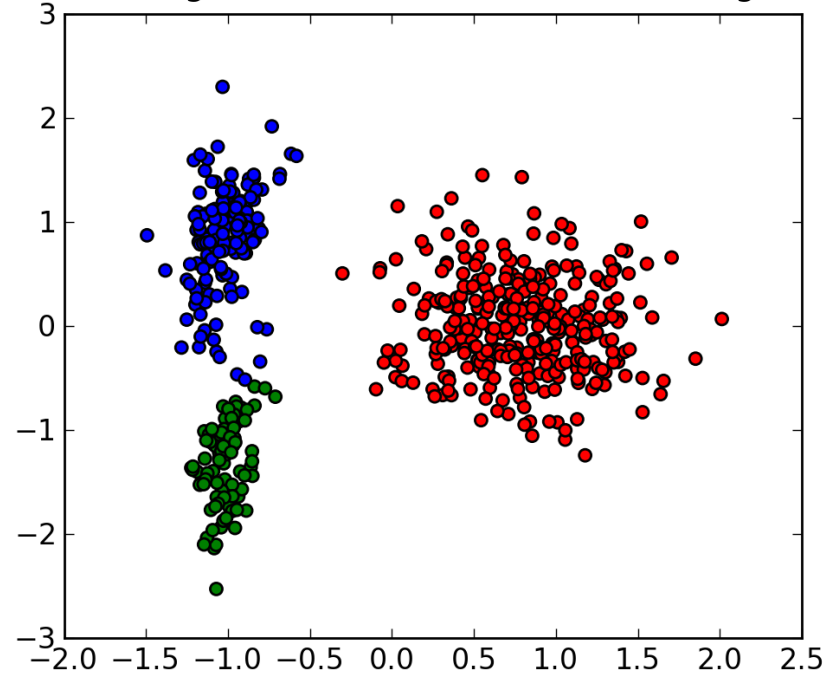




# Gráficos multivariados de las técnicas del curso

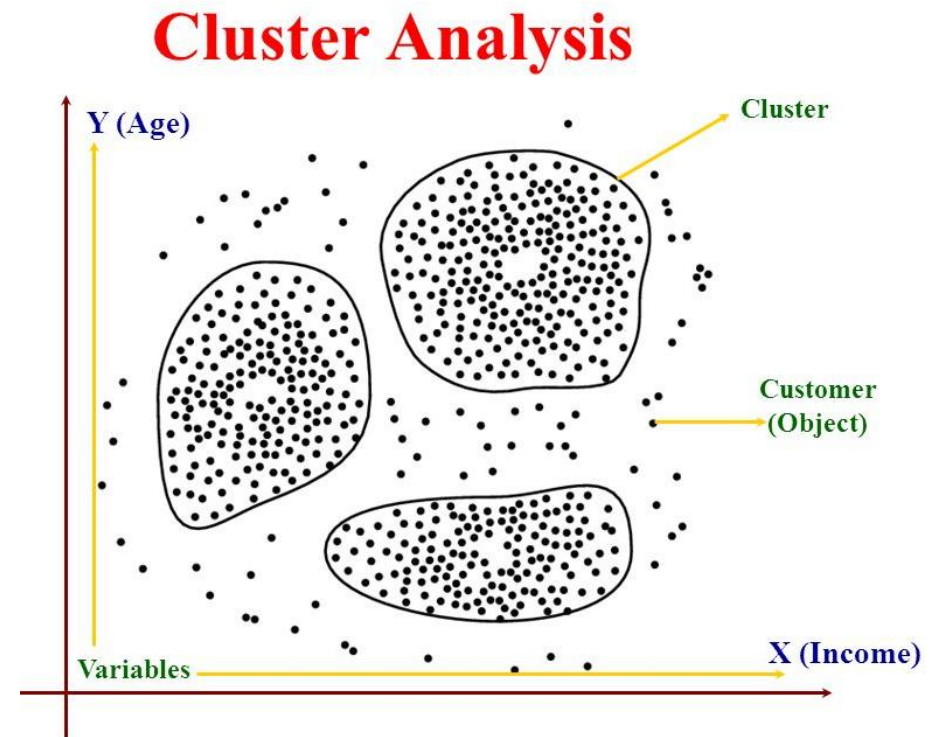
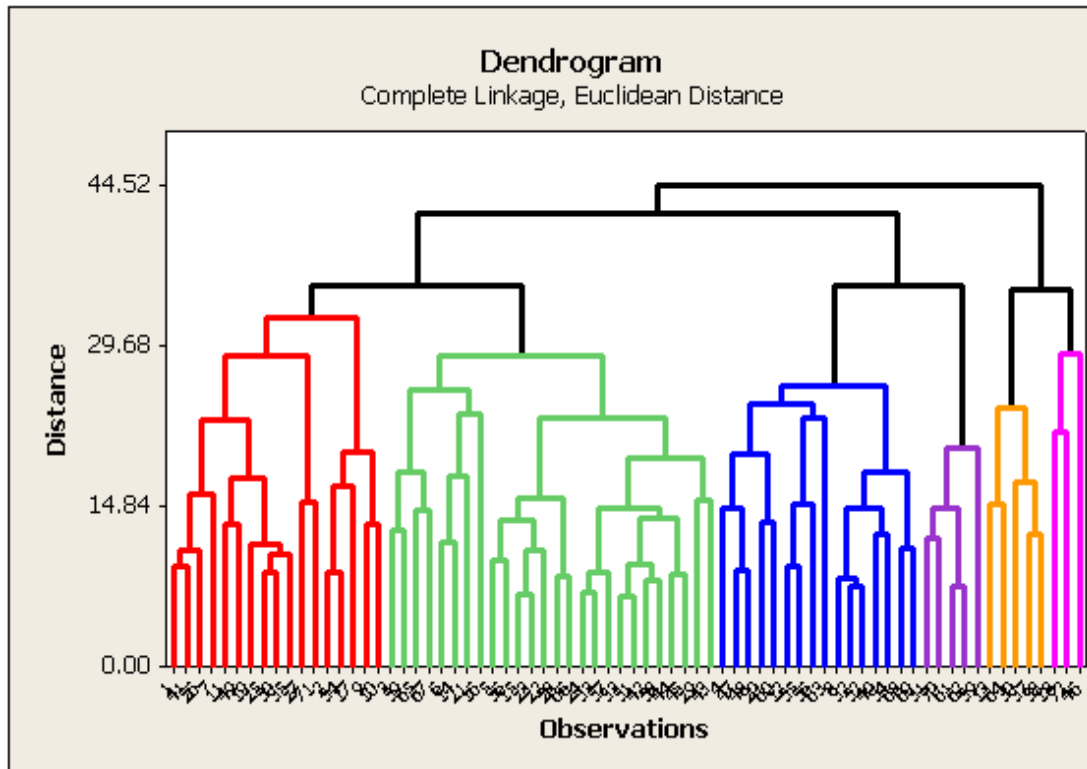
- Análisis discriminante.

Scores using FDA with multivariate linear regression



# Gráficos multivariados de las técnicas del curso

- Análisis de agrupamiento.

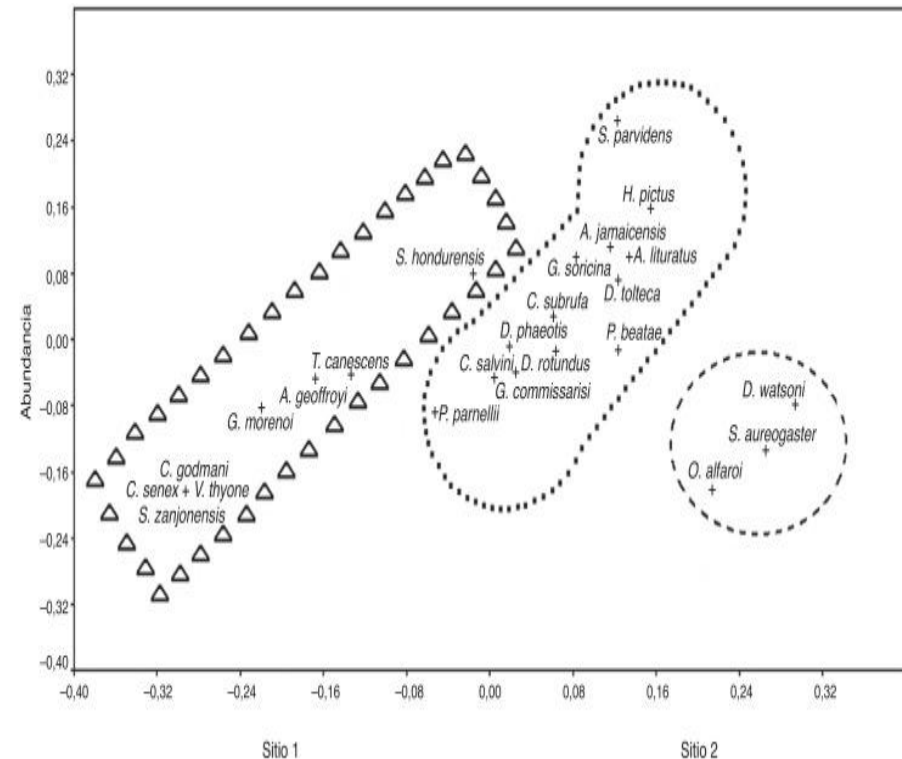
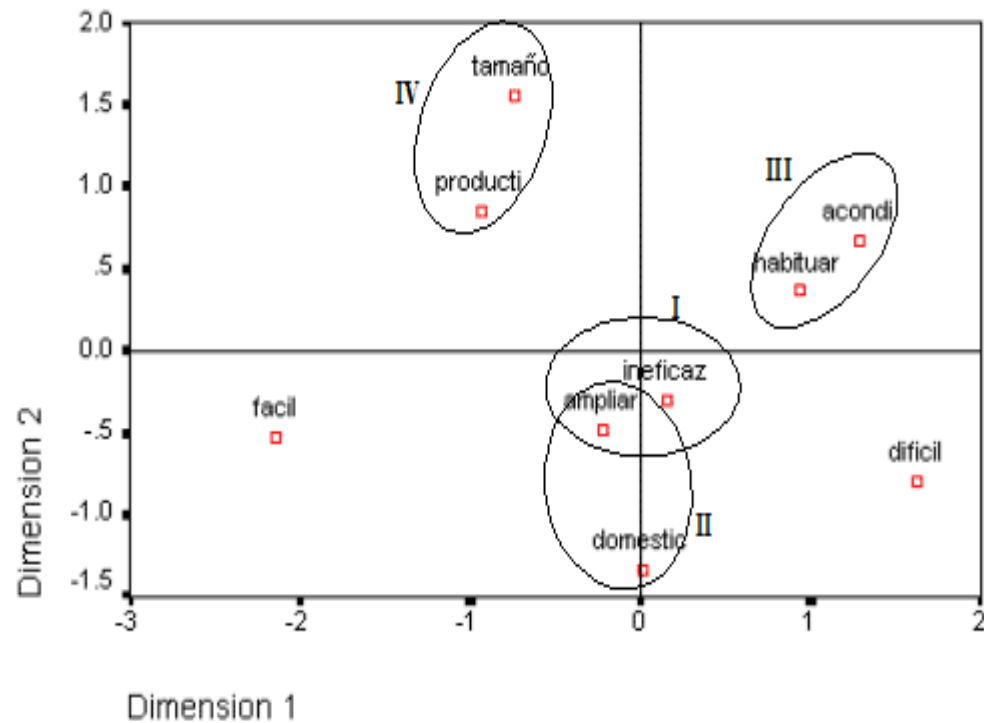


# Gráficos multivariados de las técnicas del curso

- Escalamiento multidimensional

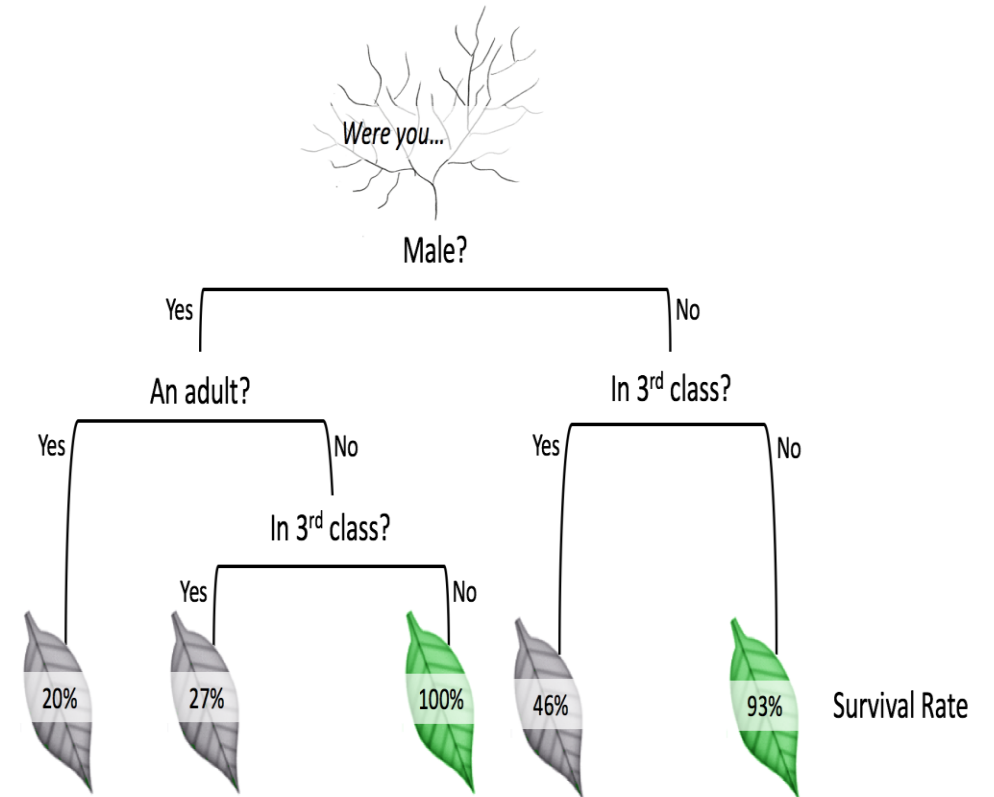
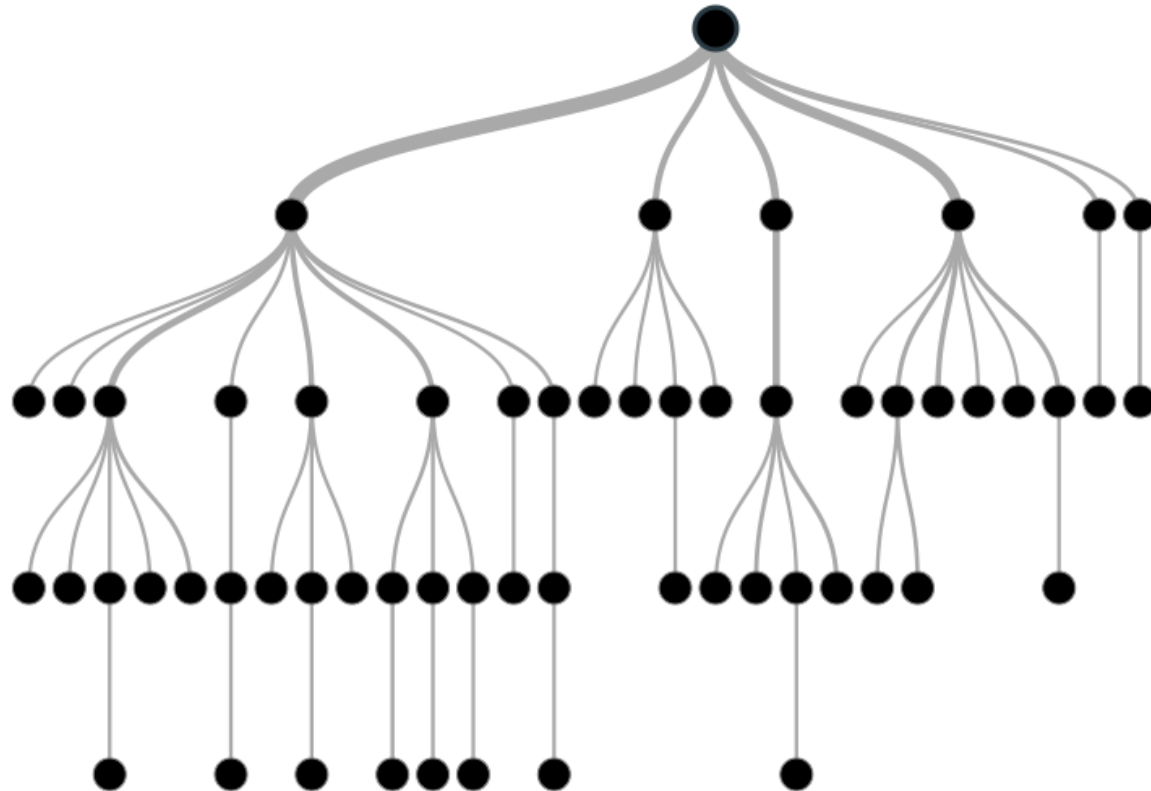
Derived Stimulus Configuration

Euclidean distance model



# Gráficos multivariados de las técnicas del curso

- Árboles de decisión



¿Cuál es la importancia de los primeros gráficos?





*The  
End*