

Usando R para visualizar los datos

Dr. Christian DiCanio

Universidad de Búfalo

I. Procedimientos básicos

- (1) El mandato básico de `plot()` produce un plot de dispersión para datos continuos, proporciones para datos de cuentas, y plots de cajas para datos continuos y discretos.
- (2) Se controla los ejes por `ylim()` y `xlim()`. Los ejes y la figura tienen etiqueta usando `ylab="texto"`, `xlab="texto"`, `main="texto."`
- (3) Se puede manipular el output por la función `par()`. Nos permite controlar el número de parámetros.
- (4) Cuando cambiamos un parámetro con `par()`, cambiará el parámetros para todas las figuras hasta que se reinicia R o hasta que se lo cambia otra vez. No existe una opción "volver a la configuración por defecto."

(5) Mandatos útiles de `par()`

<code>mfrow()</code> and <code>mfcoll()</code>	el número de columnas/filas para hacer un plot; crea figuras múltiples en un solo imagen.
<code>mar=c(5,4,4,2)</code>	controla los márgenes: al fondo, la izquierda, arriba, y a la derecha.
<code>cex=10</code>	tamaño de fuente
<code>family="serif"</code>	tipo de fuente
<code>las=0,1,2,3</code>	orientación de los ejes

(6) Dentro de un plot, se puede controlar muchos parámetros:

<code>type="p"/"l"/"b"</code>	puntos, líneas o ambos.
<code>lty=1, 2, 3, 4</code>	tipo de línea (sólido, punteado, discontinuo, etc)
<code>pch=1,2,3,4....</code>	tipo de punto (puntos, cuadros, triángulos, etc)
<code>col="green", "red", etc.</code>	color de elementos

(7) Ejemplo con datos de VOT de mixteco.

II. Procedimientos más avanzados - plotmeans()

(8) plotmeans(), una parte del paquete gplots nos permite hacer un plot de los valores que varían en tiempo con intervalos de confianza.

(9) Es útil para datos de "series de tiempo", donde se cambian los valores en el tiempo. Para hacer un plot de un serie de tiempo usando plotmeans, hay que crear una estructura de formato largo, p.ej. tiempo por valor.

(10) Los intervalos de confianza son utiles porque ilustran la varianza dentro de la distribución.

(11) Se determina el intervalo por una probabilidad basada en la distribución t. El default es $p = 0.95$, pero uno se puede ajustarlo a $p = .99$, $p = .999$ y usar la distribución normal en vez de t.

(12) Si no se quiere incluir n, cambia n.label=FALSE .

(13) Tonos de triqui en R.

III. Herramientas nuevas - ggplot2

(14) Se ha puesto el estándar para gráficos de alta calidad por su habilidad a incluir una cantidad inmensa de tipos de datos y visualizar dimensiones multiples.

(15) Se maneja los comandos básicos con la función 'qplot.' qplot() usa un sintáxis parecido a lo que se ve en plot() en R, pero puede incluir unas interacciones también.

(16) Investigamos los datos de VOT de mixteco otra vez.

(17) El sintáxis de ggplot() es muy diferente de lo que se ve en qplot y con otros comandos.

(18) Se traza una dimension X (un estético) por otra dimension Y (un estético) en términos de un objeto (geometría) con parametros (ylim, xlim, discrete, escalas continuas, etc).

(19) El diseño crea un objeto o estructura básica. Se puede agregar elementos a este objeto básico como facet_wrap() para crear figuras adicionales.

(20) Muy útil con modelos de muchos niveles.