

T501

Ing. Maximiliano Carsi Castrejón – Extracción y
Conocimiento en Bases de Datos

DESCRIPCIÓN BREVE

Este documento trata sobre la documentación extensa
y recursos para la creación de gráficos en R

Luis Eduardo Bahena Castillo

9°C IDyGS



INTRODUCCIÓN

Tarea: Explorando el Poder de ggplot2 para la Visualización de Datos en R

Objetivo: Familiarizarse con la librería ggplot2 en R y comprender sus principios fundamentales para la creación de gráficos de alta calidad para la visualización de datos.

Descripción:

En esta tarea, te sumergirás en el mundo de ggplot2, una de las librerías más populares y versátiles de R para la creación de gráficos. Investigarás sus componentes principales y aprenderás a construir visualizaciones atractivas y informativas.

Pasos a seguir:

1. Introducción a ggplot2:

- Investiga qué es ggplot2 y por qué es tan popular en la comunidad de R.
- Describe la filosofía detrás de la "gramática de gráficos" y cómo se aplica en ggplot2.
- Instala la librería ggplot2 en tu entorno de R.

2. Componentes principales de ggplot2:

- Define y explica los siguientes conceptos clave de ggplot2:
 - **Datos (data):** ¿Qué tipo de datos se pueden usar en ggplot2?
 - **Estéticas (aes):** ¿Cómo se mapean las variables a las propiedades visuales del gráfico?
 - **Geometrias (geoms):** ¿Qué tipos de gráficos se pueden crear con ggplot2 (ej: puntos, líneas, barras, etc.)?
 - **Facetas (facets):** ¿Cómo se pueden crear subgráficos para explorar diferentes subconjuntos de datos?
 - **Escalas (scales):** ¿Cómo se controlan los rangos y transformaciones de los ejes?
 - **Temas (themes):** ¿Cómo se personaliza la apariencia general del gráfico?

3. Creando gráficos básicos con ggplot2:

- Utiliza un conjunto de datos de tu elección (puedes usar uno de los conjuntos de datos incorporados en R o importar uno propio).
- Crea al menos tres gráficos diferentes utilizando ggplot2, explorando diferentes geometrias y estéticas.
- Describe el código utilizado para cada gráfico y explica las decisiones que tomaste en cuanto a la elección de geometrias, estéticas y otros parámetros.

4. Recursos adicionales:

- Investiga y lista al menos tres recursos online (tutoriales, documentación, blogs) que te hayan sido útiles para aprender sobre ggplot2.

Entregables:

- Un documento (PDF o similar) que contenga:
 - Respuestas a las preguntas de los pasos 1 y 2.
 - Código y descripción de los gráficos creados en el paso 3.
 - Lista de recursos adicionales del paso 4.

Criterios de evaluación:

- Claridad y precisión en las respuestas.
- Correcta aplicación de los conceptos de ggplot2 en la creación de gráficos.
- Calidad y legibilidad del código.
- Presentación organizada y profesional del documento.

Fuentes de Investigación

- <https://ggplot2.tidyverse.org>
- <https://es.r4ds.hadley.nz/03-visualize.html>
- <https://rpubs.com/dsulmont/37913>

DESARROLLO

Introducción a ggplot2

¿Qué es ggplot2? **ggplot2** es una librería de R diseñada para la creación de gráficos de alta calidad, ampliamente utilizada tanto en el ámbito académico como en la industria. Fue creada por Hadley Wickham y se basa en la "gramática de gráficos", un enfoque que permite construir gráficos de manera intuitiva y estructurada. La popularidad de **ggplot2** se debe a su flexibilidad, capacidad para manejar datos complejos, y su integración con el ecosistema de R.

Filosofía de la "Gramática de Gráficos"

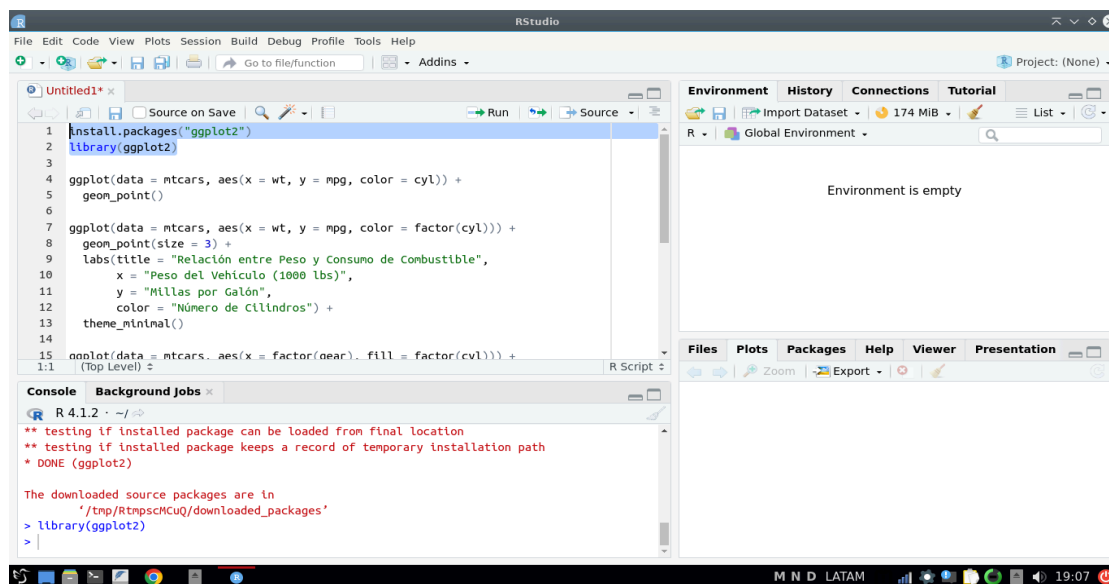
La "gramática de gráficos" es el principio fundamental detrás de **ggplot2**. Esta filosofía postula que cualquier gráfico puede descomponerse en una serie de componentes básicos: datos, estéticas, geometrías, escalas, facetas y temas. Al combinar estos elementos, es posible construir una amplia variedad de gráficos. Esta estructura modular permite a los usuarios personalizar y expandir sus gráficos fácilmente, lo que hace de **ggplot2** una herramienta poderosa y versátil para la visualización de datos.

Instalación de ggplot2

Para comenzar a utilizar **ggplot2**, es necesario instalar la librería en el entorno de R. Esto se realiza mediante el siguiente comando:

```
install.packages("ggplot2")
```

```
library(ggplot2)
```



The screenshot shows the RStudio environment. The script editor on the left contains the following code:

```
1 install.packages("ggplot2")
2 library(ggplot2)
3
4 ggplot(data = mtcars, aes(x = wt, y = mpg, color = cyl)) +
5   geom_point()
6
7 ggplot(data = mtcars, aes(x = wt, y = mpg, color = factor(cyl))) +
8   geom_point(size = 3) +
9   labs(title = "Relación entre Peso y Consumo de Combustible",
10        x = "Peso del Vehículo (1000 lbs)",
11        y = "Millas por Galón",
12        color = "Número de Cilindros") +
13   theme_minimal()
14
15 #ggplot(data = mtcars, aes(x = factor(gear), fill = factor(cyl))) +
```

The console on the bottom left shows the output of the installation process:

```
R 4.1.2 ~ /
** testing if installed package can be loaded from final location
** testing if installed package keeps a record of temporary installation path
* DONE (ggplot2)

The downloaded source packages are in
  '/tmp/RtmpscMCuQ/downloaded_packages'
> library(ggplot2)
>
```

The Environment pane on the right shows "Global Environment" and "Environment is empty". The bottom status bar indicates the system is running on a Linux machine (M N D LATAM) at 19:07.

Componentes Principales de ggplot2

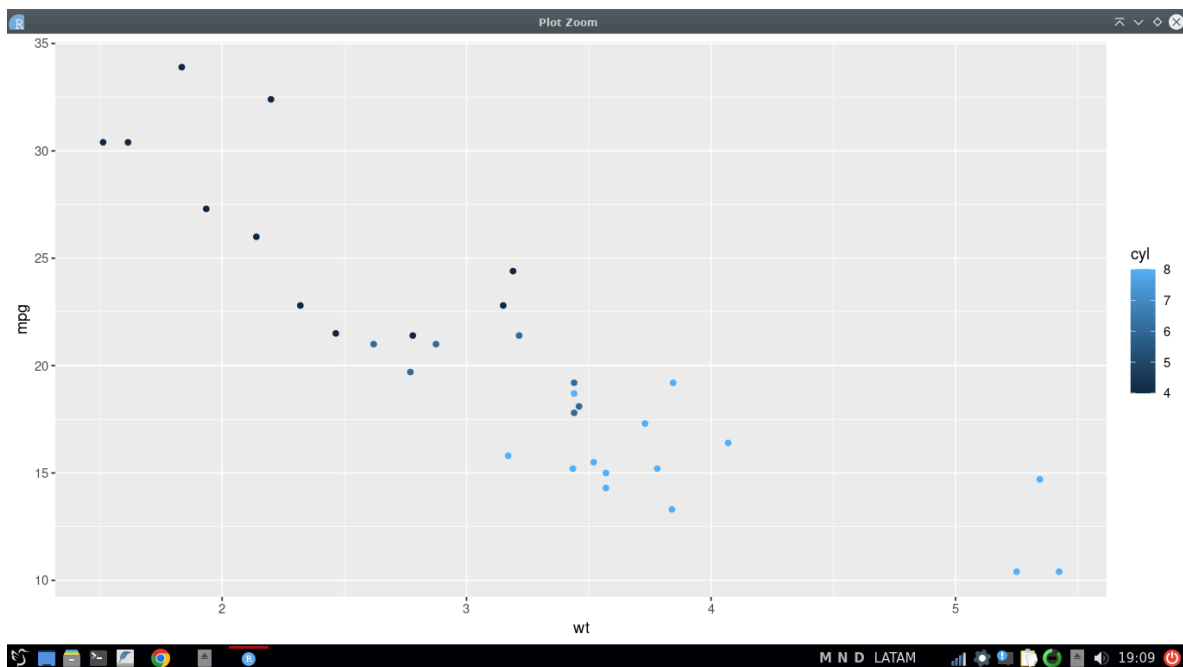
Datos (Data)

En **ggplot2**, los datos son el punto de partida para cualquier gráfico. Se pueden usar diferentes tipos de datos, como data frames, tibbles o matrices, siempre que las variables estén organizadas en columnas. **ggplot2** permite trabajar con datos categóricos, numéricos y de fecha, y se adapta bien a conjuntos de datos grandes o complejos.

Estéticas (aes)

Las estéticas en **ggplot2** se refieren a cómo se mapean las variables del conjunto de datos a propiedades visuales del gráfico, como posición, color, tamaño y forma. Por ejemplo, en un gráfico de dispersión, las variables **x** e **y** pueden representarse mediante las coordenadas del gráfico, mientras que una tercera variable puede mapearse al color o tamaño de los puntos:

```
ggplot(data = mtcars, aes(x = wt, y = mpg, color = cyl)) +  
  geom_point()
```



Geometrías (geoms)

Las geometrías son los componentes visuales del gráfico, como puntos, líneas o barras. Cada tipo de gráfico en `ggplot2` se construye a partir de una o más geometrías. Por ejemplo, `geom_point()` crea un gráfico de dispersión, `geom_line()` crea un gráfico de líneas, y `geom_bar()` se utiliza para gráficos de barras. La elección de geometría depende del tipo de datos y del mensaje que se quiere transmitir.

Facetas (facets)

Las facetas permiten dividir un gráfico en subgráficos basados en los valores de una o más variables. Esto es útil para explorar la relación entre variables en diferentes subconjuntos de datos. `ggplot2` proporciona dos funciones principales para crear facetas: `facet_wrap()` y `facet_grid()`, que permiten crear subgráficos de manera flexible.

Escalas (scales)

Las escalas en `ggplot2` controlan cómo se mapean las estéticas a las propiedades visuales. Esto incluye la transformación de ejes, la definición de límites y la personalización de etiquetas. Las escalas permiten ajustar la visualización de los datos para mejorar la interpretación y la presentación del gráfico.

Temas (themes)

Los temas en `ggplot2` permiten personalizar la apariencia general del gráfico, incluyendo elementos como el fondo, las líneas de cuadrícula, las etiquetas de los ejes y la leyenda. `ggplot2` incluye varios temas prediseñados, como `theme_bw()` o `theme_minimal()`, y permite la creación de temas personalizados para adaptarse a las necesidades estéticas del usuario.

Creando Gráficos Básicos con ggplot2

Para ilustrar el uso de `ggplot2`, se presentan tres ejemplos de gráficos utilizando el conjunto de datos `mtcars`, que viene incorporado en R.

Gráfico de Dispersión

Este gráfico de dispersión muestra la relación entre el peso del vehículo (`wt`) y el consumo de combustible (`mpg`), diferenciando los puntos por el número de cilindros (`cyl`). Se optó por un tema minimalista para resaltar los puntos y facilitar la lectura del gráfico.

```

ggplot(data = mtcars, aes(x = wt, y = mpg, color = factor(cyl))) +

  geom_point(size = 3) +

  labs(title = "Relación entre Peso y Consumo de Combustible",

        x = "Peso del Vehículo (1000 lbs)",

        y = "Millas por Galón",

        color = "Número de Cilindros") +

  theme_minimal()
  
```

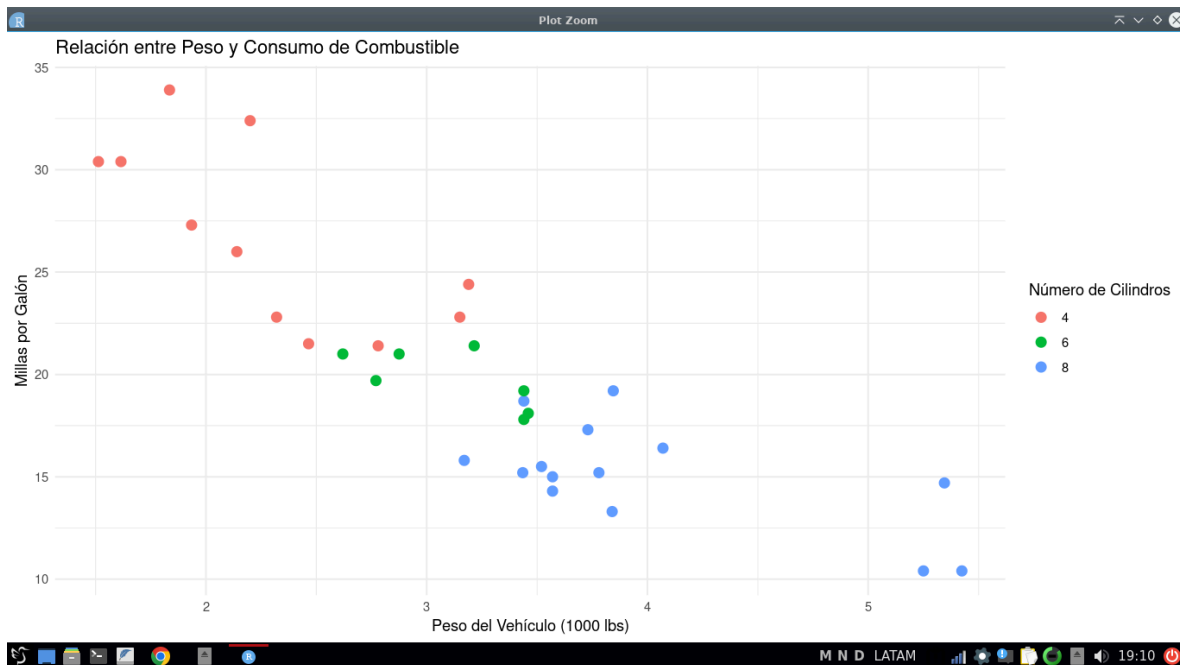


Gráfico de Barras

Este gráfico de barras compara la distribución de vehículos según el número de marchas (**gear**) y cilindros (**cyl**). Se utilizó **geom_bar()** con la opción **position = "dodge"** para separar las barras por cilindros, facilitando la comparación directa entre categorías.

```

ggplot(data = mtcars, aes(x = factor(gear), fill = factor(cyl))) +

  geom_bar(position = "dodge") +

  labs(title = "Distribución de Vehículos por Número de Marchas",

        x = "Número de Marchas",

        y = "Cuenta de Vehículos",
  
```

```
fill = "Número de Cilindros") +
```

```
theme_bw()
```

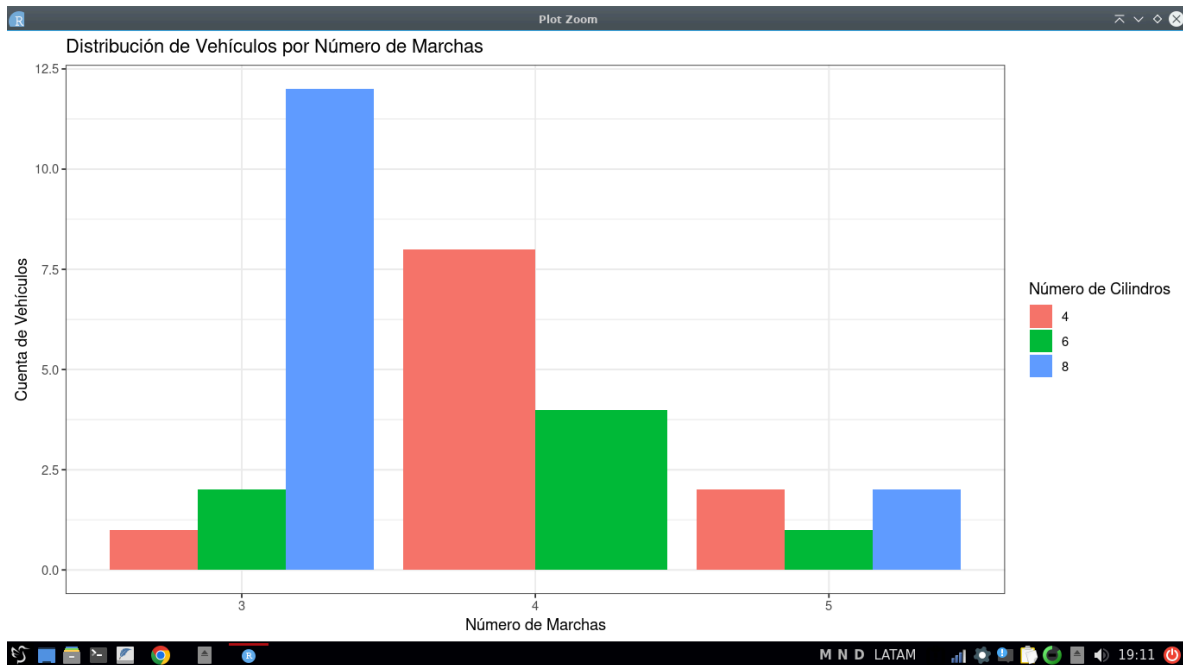


Gráfico de Líneas

Este gráfico de líneas muestra la evolución del número de desempleados en EE.UU. a lo largo del tiempo, utilizando el conjunto de datos `economics`. Se eligió un tema clásico y una línea azul para dar un aspecto profesional y claro.

```
ggplot(data = economics, aes(x = date, y = unemploy)) +
```

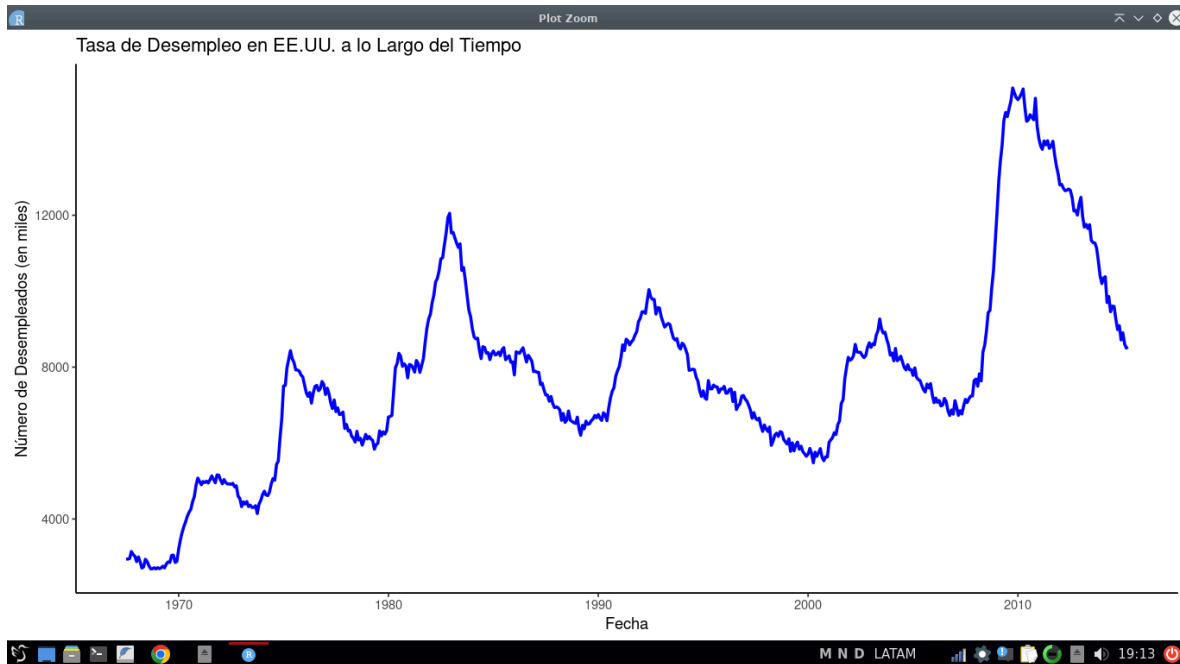
```
  geom_line(color = "blue", size = 1) +
```

```
  labs(title = "Tasa de Desempleo en EE.UU. a lo Largo del Tiempo",
```

```
        x = "Fecha",
```

```
        y = "Número de Desempleados (en miles)") +
```

```
  theme_classic()
```

CONCLUSIÓN

El poder de **ggplot2** radica en su capacidad para transformar datos complejos en visualizaciones claras y efectivas, facilitando la comprensión y el análisis de la información. La "gramática de gráficos" que sustenta a **ggplot2** proporciona una estructura modular y flexible, que permite a los usuarios construir gráficos de manera sistemática y personalizable. Esta metodología no solo mejora la precisión en la representación de datos, sino que también ofrece una gran libertad creativa en la presentación visual.

Al comprender y aplicar los principios fundamentales de **ggplot2**, como la manipulación de datos, el mapeo de estéticas, la selección de geometrías, y la personalización de temas, los usuarios pueden diseñar gráficos que no solo comunican los patrones y tendencias de manera efectiva, sino que también están adaptados a las necesidades específicas de su audiencia. Por ejemplo, la capacidad para crear facetas permite explorar subconjuntos de datos en profundidad, mientras que las escalas y temas personalizables aseguran que los gráficos sean tanto informativos como visualmente atractivos.

La flexibilidad de **ggplot2** también se manifiesta en su capacidad para integrarse con otros paquetes y herramientas en el ecosistema de R, ampliando las posibilidades de análisis y visualización. La posibilidad de combinar **ggplot2** con paquetes como **dplyr** para la manipulación de datos o **shiny** para aplicaciones web interactivas, convierte a **ggplot2** en una herramienta indispensable para analistas y científicos de datos.