Clase_2 DataSaurus 2025 con compilador R Markdown

mateo

2025-04-03

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

Librerias necesarias

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4      v readr     2.1.5
## v forcats   1.0.0      v stringr   1.5.1
## v ggplot2   3.5.1      v tibble    3.2.1
## v lubridate 1.9.4      v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Dataset para el análisis de datos
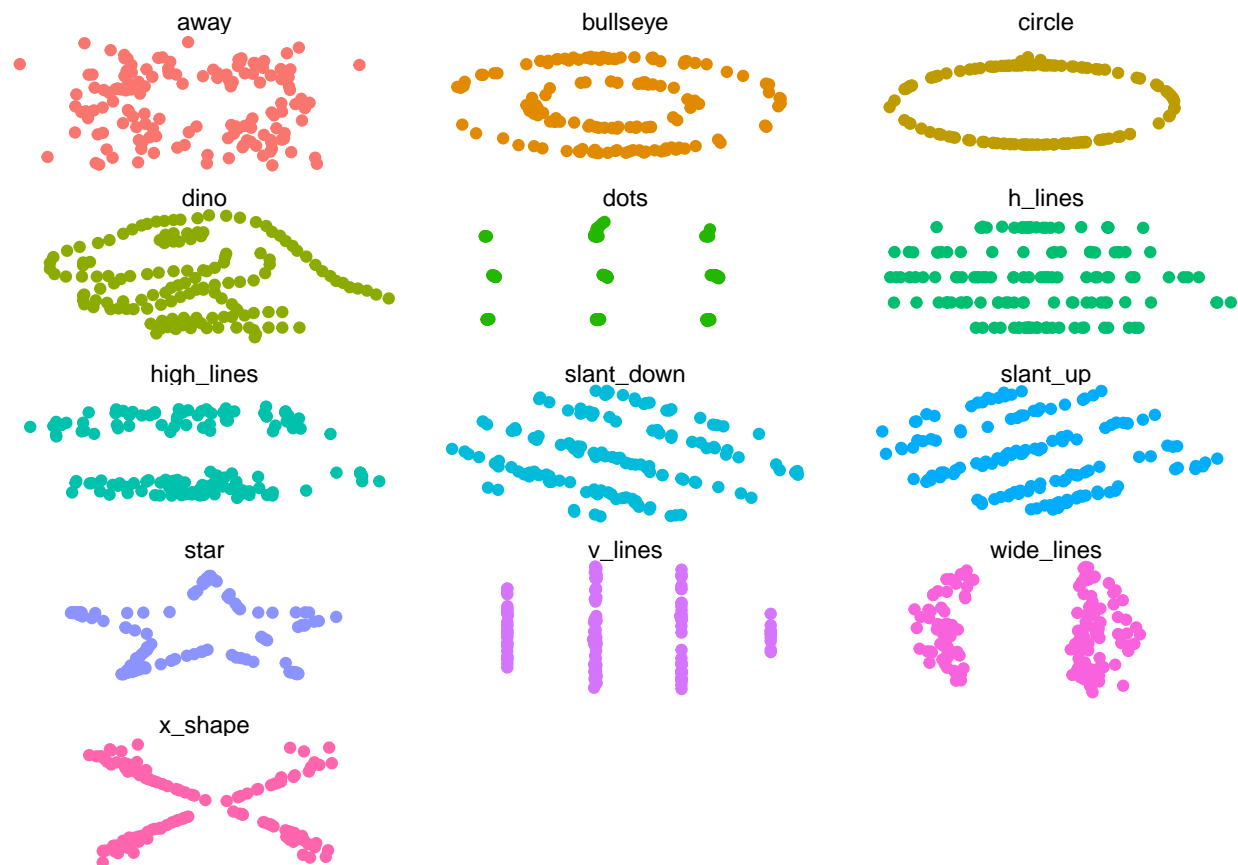
```r
datasaurus_dozen %>%
  group_by(dataset) %>%
  summarize(
    mean_x = mean(x),
    mean_y = mean(y),
    std_dev_x = sd(x),
    std_dev_y = sd(y),
    corr_x_y = cor(x,y)
  )
```

```
## # A tibble: 13 x 6
##    dataset     mean_x mean_y std_dev_x std_dev_y corr_x_y
##    <chr>        <dbl>  <dbl>     <dbl>     <dbl>    <dbl>
## 1 away          54.3   47.8      16.8      26.9  -0.0641
## 2 bullseye      54.3   47.8      16.8      26.9  -0.0686
## 3 circle        54.3   47.8      16.8      26.9  -0.0683
## 4 dino          54.3   47.8      16.8      26.9  -0.0645
## 5 dots          54.3   47.8      16.8      26.9  -0.0603
## 6 h_lines       54.3   47.8      16.8      26.9  -0.0617
## 7 high_lines    54.3   47.8      16.8      26.9  -0.0685
## 8 slant_down    54.3   47.8      16.8      26.9  -0.0690
```

```
##  9 slant_up      54.3    47.8       16.8       26.9   -0.0686
## 10 star          54.3    47.8       16.8       26.9   -0.0630
## 11 v_lines       54.3    47.8       16.8       26.9   -0.0694
## 12 wide_lines    54.3    47.8       16.8       26.9   -0.0666
## 13 x_shape       54.3    47.8       16.8       26.9   -0.0656
```

Gráfica generada

```r
ggplot(datasaurus_dozen, aes(x=x, y=y, colour = dataset))+
  geom_point()+
  theme_void()+
  theme(legend.position = "none")+
  facet_wrap(~dataset, ncol = 3)
```



```r
# Cálculos estadísticos opcionales
datasaurus_dozen %>%
  group_by(dataset) %>%
  summarize(
    mean_x = mean(x),
    mean_y = mean(y),
    std_dev_x = sd(x),
    std_dev_y = sd(y),
    corr_x_y = cor(x,y)
  )
```

```
## # A tibble: 13 x 6
##    dataset     mean_x mean_y std_dev_x std_dev_y corr_x_y
```

```
##    <chr>        <dbl>  <dbl>   <dbl>    <dbl>    <dbl>
## 1  away          54.3   47.8    16.8     26.9  -0.0641
## 2  bullseye      54.3   47.8    16.8     26.9  -0.0686
## 3  circle        54.3   47.8    16.8     26.9  -0.0683
## 4  dino          54.3   47.8    16.8     26.9  -0.0645
## 5  dots          54.3   47.8    16.8     26.9  -0.0603
## 6  h_lines       54.3   47.8    16.8     26.9  -0.0617
## 7  high_lines    54.3   47.8    16.8     26.9  -0.0685
## 8  slant_down    54.3   47.8    16.8     26.9  -0.0690
## 9  slant_up      54.3   47.8    16.8     26.9  -0.0686
## 10 star          54.3   47.8    16.8     26.9  -0.0630
## 11 v_lines       54.3   47.8    16.8     26.9  -0.0694
## 12 wide_lines    54.3   47.8    16.8     26.9  -0.0666
## 13 x_shape       54.3   47.8    16.8     26.9  -0.0656
```

```r
# Gráfico de regresión lineal solo para 'slant_down'
datasaurus_dozen %>%
  filter(dataset == "slant_down") %>%
  ggplot(aes(x = x, y = y)) +
  geom_point(color = "dodgerblue") +
  geom_smooth(method = "lm", se = FALSE, color = "red", linewidth = 1) +
  labs(title = "Regresión lineal - slant_down",
       x = "x",
       y = "y") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```


Regresión lineal – slant_down