Tarea 2 Greg-Torres

Greg Torres

2024-02-08

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.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
airpolution_b <- read.csv(file = "airpollution.csv",header = TRUE, sep = ",")

library(DataExplorer)
#view(airpolution_b)
create_report(airpolution_b)

##
##
## processing file: report.rmd

##   |                                                      |
|   0%  |                                                    |.
|   2%                                                  |
|..                                                 |   5% [global_options]
|                                                       |...
|   7%                                                  |
|....                                               |  10% [introduce]
|                                                       |....
|  12%                                                  |
|.....                                              |  14% [plot_intro]
|                                                       |......
|  17%                                                  |
|.......                                            |  19% [data_structure]
|                                                       |........
|  21%                                                  |
|.........                                          |  24% [missing_profile]
|                                                       |..........
|  26%                                                  |
|...........                                        |  29% [univariate_distribution_header]
|                                                       |...........
|  31%                                                  |
|............                                       |  33% [plot_histogram]
```

```
|                                                  |.............
|  36%                                             |
|..............                                    |  38% [plot_density]
|                                                  |.............
|  40%                                             |
|..............                                    |  43% [plot_frequency_bar]
|                                                  |.............
|  45%                                             |
|..............                                    |  48% [plot_response_bar]
|                                                  |.............
|  50%                                             |
|..............                                    |  52% [plot_with_bar]
|                                                  |.............
|  55%                                             |
|..............                                    |  57% [plot_normal_qq]
|                                                  |.............
|  60%                                             |
|..............                                    |  62% [plot_response_qq]
|                                                  |.............
|  64%                                             |
|..............                                    |  67% [plot_by_qq]
|                                                  |.............
|  69%                                             |
|..............                                    |  71% [correlation_analysis]
|                                                  |.............
|  74%                                             |
|..............                                    |  76% [principal_component_analysis]
|                                                  |.............
|  79%                                             |
|..............                                    |  81% [bivariate_distribution_header]
|
|..............                                    |  83%
|
|..............                                    |  86% [plot_response_boxplot]
|
|..............                                    |  88%
|
|..............                                    |  90% [plot_by_boxplot]
|
|..............                                    |  93%
|
|..............                                    |  95% [plot_response_scatterplot]
|
|..............                                    |  98%
|
|..............                                    | 100% [plot_by_scatterplot]

## output file: C:/WorkR/report.knit.md
```

```
## "C:/Program Files/RStudio/resources/app/bin/quarto/bin/tools/pandoc" +RTS
-K512m -RTS "C:\WorkR\report.knit.md" --to html4 --from
markdown+autolink_bare_uris+tex_math_single_backslash --output
pandoc625c643377f6.html --lua-filter "C:\Users\gregt\AppData\Local\R\win-
library\4.3\rmarkdown\rmarkdown\lua\pagebreak.lua" --lua-filter
"C:\Users\gregt\AppData\Local\R\win-
library\4.3\rmarkdown\rmarkdown\lua\latex-div.lua" --embed-resources --
standalone --variable bs3=TRUE --section-divs --table-of-contents --toc-depth
6 --template "C:\Users\gregt\AppData\Local\R\win-
library\4.3\rmarkdown\rmd\h\default.html" --no-highlight --variable
highlightjs=1 --variable theme=yeti --mathjax --variable "mathjax-
url=https://mathjax.rstudio.com/latest/MathJax.js?config=TeX-AMS-
MML_HTMLorMML" --include-in-header
"C:\Users\gregt\AppData\Local\Temp\RtmpCoNqId\rmarkdown-str625c550520a2.html"

##
## Output created: report.html

colnames(airpolution_b)

## [1] "Nitrogen.Oxides"      "Respirable.Particles"

# Nitrogen.Oxides      Respirable.Particles
attach(airpolution_b)
sink("ReporteAire_b.txt")
# media
mean(Nitrogen.Oxides, na.rm = TRUE)

## [1] 187.9584

# mediana
median(Nitrogen.Oxides, na.rm = TRUE)

## [1] 164.65

min(Nitrogen.Oxides, na.rm = TRUE)

## [1] 50.9

max(Nitrogen.Oxides, na.rm = TRUE)

## [1] 587.5

fivenum(Nitrogen.Oxides, na.rm = TRUE)

## [1]  50.90 125.60 164.65 232.00 587.50

sd(Nitrogen.Oxides, na.rm = TRUE)

## [1] 81.89953

var(Nitrogen.Oxides, na.rm = TRUE)

## [1] 6707.533
```

```r
quantile(Nitrogen.Oxides, probs = 0.5,
         na.rm = T)
```

```
##    50%
## 164.65
```

```r
summary(airpolution_b)
```

```
##  Nitrogen.Oxides Respirable.Particles
##  Min.   : 50.9   Min.   :12.20
##  1st Qu.:125.7   1st Qu.:24.90
##  Median :164.7   Median :30.05
##  Mean   :188.0   Mean   :32.05
##  3rd Qu.:231.6   3rd Qu.:37.23
##  Max.   :587.5   Max.   :77.90
##  NA's   :4       NA's   :6
```

```r
sink()

levantadores <- read.csv("Halterofilia.csv", sep = ";")

Peces2 <- read.table("datos.txt", sep = "\t",
  header = TRUE)
Peces2 <- Peces2[ , -1]
Peces3 <- Peces2[25:75, c(1,2)]

library(ggplot2)
Arbol <- Orange
Arbol <- data.frame(Arbol)
colnames(Arbol)
```

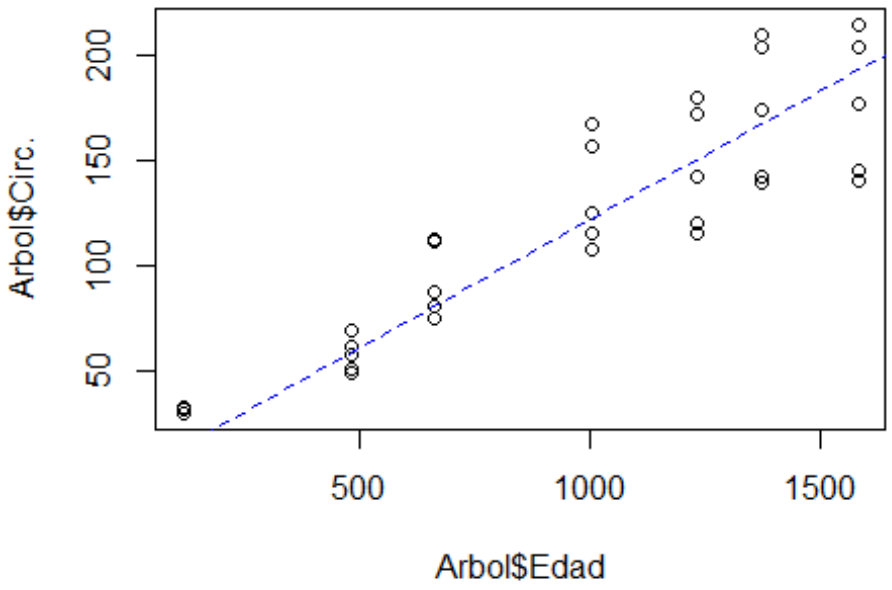
```
## [1] "Tree"          "age"           "circumference"
```

```r
Arbol <- Arbol[ , -1]
names(Arbol) <- c("Edad", "Circ.")
mod2 <- lm(Circ. ~ Edad -1, data = Arbol)
summary(mod2)
```

```
##
## Call:
## lm(formula = Circ. ~ Edad - 1, data = Arbol)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -52.297  -8.832   6.289  20.180  44.960
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## Edad 0.121553   0.004022   30.23   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 24.79 on 34 degrees of freedom
## Multiple R-squared:  0.9641, Adjusted R-squared:  0.9631
## F-statistic: 913.6 on 1 and 34 DF,  p-value: < 2.2e-16
```

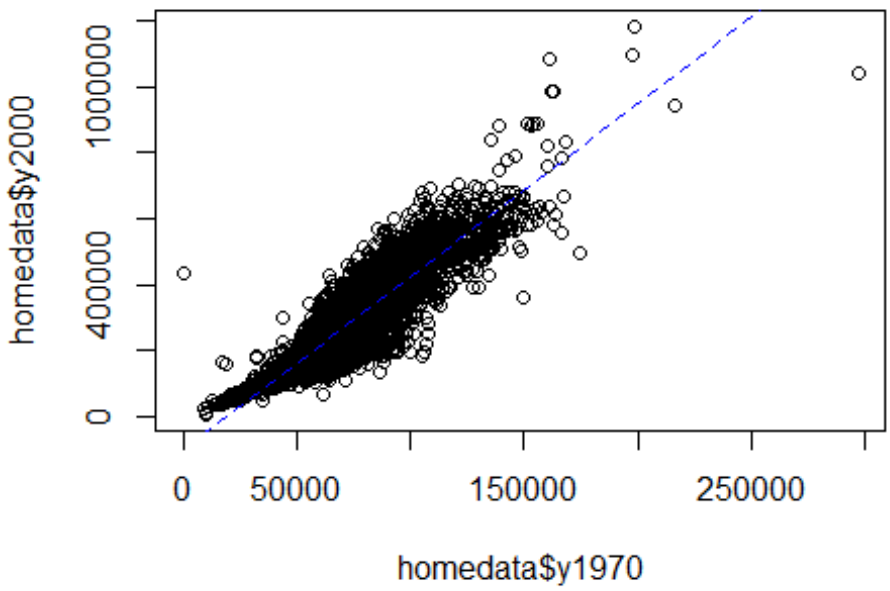
```
plot(x = Arbol$Edad, y = Arbol$Circ.)
abline(mod2, col="blue", lty=2)
```




```
library(report)

report(mod2)
```

```
## We fitted a linear model (estimated using OLS) to predict Circ. with Edad
## (formula: Circ. ~ Edad - 1). The model explains a statistically
significant and
## substantial proportion of variance (R2 = 0.96, F(1, 34) = 913.56, p <
.001,
## adj. R2 = 0.96). The model's intercept, corresponding to Edad = 0, is at
(t() =
## , p ). Within this model:
##
##   - The effect of Edad is statistically significant and positive (beta =
0.12,
## 95% CI [0.11, 0.13], t(34) = 30.23, p < .001; Std. beta = 0.91, 95% CI
[0.77,
## 1.06])
##
```

```
## Standardized parameters were obtained by fitting the model on a
standardized
## version of the dataset. 95% Confidence Intervals (CIs) and p-values were
## computed using a Wald t-distribution approximation.

x = Arbol$Edad
 y = 0.12*x
 0.13*800

## [1] 104

library(UsingR)

## Loading required package: MASS

## Loading required package: HistData

## Loading required package: Hmisc

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units

data("homedata")
head(homedata)

##    y1970  y2000
## 1  89700 359100
## 2 118400 504500
## 3 116400 477300
## 4 122000 500400
## 5  91500 433900
## 6 102800 464800

tail(homedata)

##       y1970  y2000
## 6836  58100 107400
## 6837 105500 183100
## 6838  87000 135600
## 6839  35200  52100
## 6840  10000  12700
## 6841  61700  69800

plot(homedata$y2000 ~ homedata$y1970)
mod3 <- lm(y2000 ~ y1970, data=homedata)
abline(mod3, col="blue", lty=2)
```




```
summary(mod3)

##
## Call:
## lm(formula = y2000 ~ y1970, data = homedata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -416665  -36308     809   34372  536605
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.040e+05  2.337e+03  -44.51   <2e-16 ***
## y1970        5.258e+00  3.147e-02  167.07   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 58000 on 6839 degrees of freedom
## Multiple R-squared:  0.8032, Adjusted R-squared:  0.8032
## F-statistic: 2.791e+04 on 1 and 6839 DF,  p-value: < 2.2e-16

report(mod3)

## We fitted a linear model (estimated using OLS) to predict y2000 with y1970
## (formula: y2000 ~ y1970). The model explains a statistically significant
and
## substantial proportion of variance (R2 = 0.80, F(1, 6839) = 27912.40, p <
```

```
.001,
## adj. R2 = 0.80). The model's intercept, corresponding to y1970 = 0, is at
## -1.04e+05 (95% CI [-1.09e+05, -99424.32], t(6839) = -44.51, p < .001).
Within
## this model:
##
##   - The effect of y1970 is statistically significant and positive (beta =
5.26,
## 95% CI [5.20, 5.32], t(6839) = 167.07, p < .001; Std. beta = 0.90, 95% CI
## [0.89, 0.91])
##
## Standardized parameters were obtained by fitting the model on a
standardized
## version of the dataset. 95% Confidence Intervals (CIs) and p-values were
## computed using a Wald t-distribution approximation.
```

```
ggplot(homedata) +
  aes(x = y1970, y = y2000) +
  geom_point() +
  labs(x= "Año 1970", y="Año 2000") +
  geom_smooth(method = lm, show.legend = T) +
  theme_minimal()
```

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