R Markdown Tutorial

Ruben Bach and Christoph Kern

2024-09-11

Contents

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:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Same Code, but results only.

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Still same code but this time code only and no results!

```
summary(cars)
```

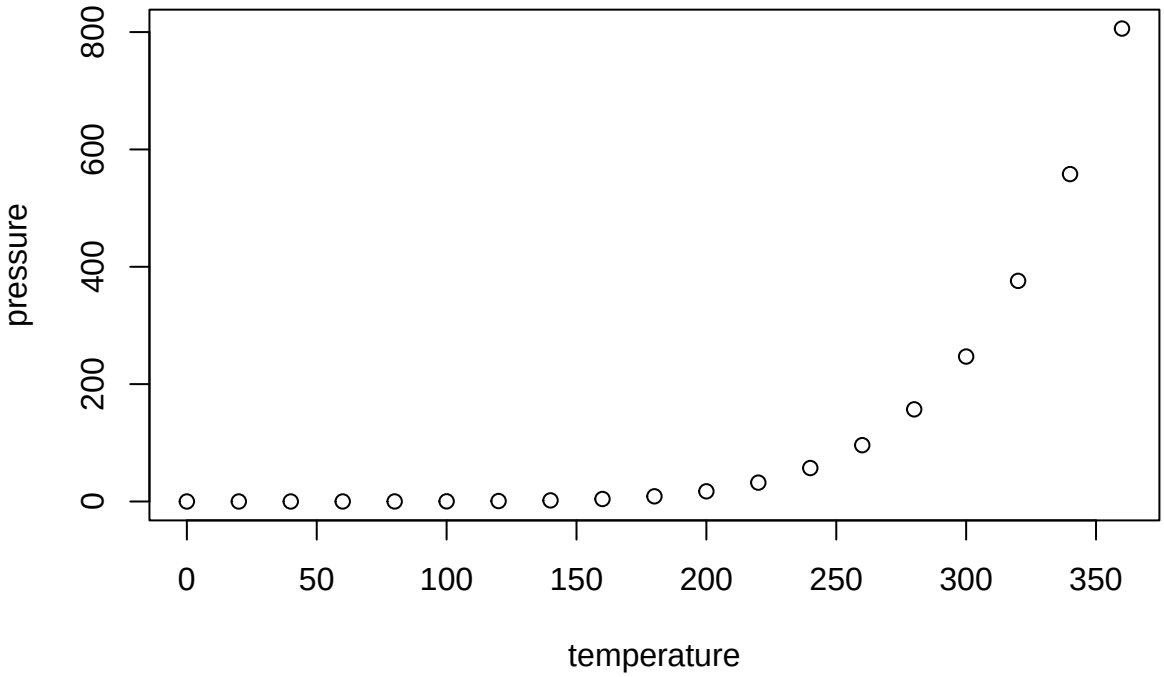
Data tables can be formatted in different styles.

```
cars[1:5, ]
```

```
##   speed dist
## 1     4    2
## 2     4   10
## 3     7    4
## 4     7   22
## 5     8   16
```

Including Plots

You can also embed plots, for example:



For headers, use #, ##, ###, ####, #####, ###### t

A large header

A slightly smaller header with a bold word (two asterisks)

For *italic* font use one asterisk. We can also ~~cross out~~ words!
Two begin a new line, add two spaces to the end of a line.
A new line starts!

Citations

Seminal work on propensity score (Rosenbaum and Rubin 1983).
Year cited without author (1983).
Rosenbaum and Rubin (1983) cited in line.

Installing scripts

Installing packages

We often need to include the repository from where a package should be installed. Also, it is good practice to check whether a required package is already installed and then install it only if it isn't.

```
if (!require('MASS')) install.packages('MASS', repos = "http://cran.us.r-project.org")
library('MASS')
if (!require('ggplot2')) install.packages('ggplot2', repos = "http://cran.us.r-project.org")
library('ggplot2')
if (!require('dplyr')) install.packages('dplyr', repos = "http://cran.us.r-project.org")
library('dplyr')
```

Example analysis

We estimate a linear regression model explaining the price of a diamond by its clarity, cut, and carat. By the way, the mean of c(1:5) is 3.

```
regmod.1 <- lm(price ~ clarity + cut + carat, data = diamonds)
summary(regmod.1)
```

```
##
## Call:
## lm(formula = price ~ clarity + cut + carat, data = diamonds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16842.5   -636.4   -114.3    474.8  11238.6
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept) -3187.540     14.475 -220.208   <2e-16 ***
## clarity.L    4011.681     33.931  118.231   <2e-16 ***
## clarity.Q   -1821.922     31.870  -57.167   <2e-16 ***
## clarity.C     917.658     27.313   33.598   <2e-16 ***
## clarity^4    -430.047     21.831  -19.699   <2e-16 ***
## clarity^5     257.141     17.821   14.429   <2e-16 ***
## clarity^6      26.909     15.539    1.732   0.0833 .
## clarity^7     186.742     13.685   13.646   <2e-16 ***
## cut.L         713.804     22.511   31.709   <2e-16 ***
## cut.Q        -334.503     19.828  -16.871   <2e-16 ***
## cut.C         188.482     17.218   10.947   <2e-16 ***
## cut^4           1.663     13.794    0.121   0.9040
## carat        8472.026     12.615  671.584   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1281 on 53927 degrees of freedom
## Multiple R-squared:  0.8969, Adjusted R-squared:  0.8969
## F-statistic: 3.911e+04 on 12 and 53927 DF,  p-value: < 2.2e-16
```

Parameters

```
print(paste("The years is ", params$year))
```

```
## [1] "The years is  2020"
```

```
regmod.2 <- lm(mpg ~ cyl + disp, data = params$data)
summary(regmod.2)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + disp, data = params$data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4213 -2.1722 -0.6362  1.1899  7.0516
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.66099    2.54700  13.609 4.02e-14 ***
## cyl         -1.58728    0.71184  -2.230   0.0337 *
## disp        -0.02058    0.01026  -2.007   0.0542 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.055 on 29 degrees of freedom
## Multiple R-squared:  0.7596, Adjusted R-squared:  0.743
## F-statistic: 45.81 on 2 and 29 DF,  p-value: 1.058e-09
```

For a nicer look of the regression output, we can use the stargazer package

```
library(stargazer)
stargazer(regmod.1,
          title = "Regression table with stargazer",
          label="tab1",
          model.numbers = FALSE,
          type="latex")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Wed, Sep 11, 2024 - 03:16:20 PM

Simply include a references heading at the end. References will be added automatically.

References

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1):41–55.

Table 1: Regression table with stargazer

	Dependent variable:
	price
clarity.L	4,011.681***
	(33.931)
clarity.Q	−1,821.922***
	(31.870)
clarity.C	917.658***
	(27.313)
clarity^4	−430.047***
	(21.831)
clarity^5	257.141***
	(17.821)
clarity^6	26.909*
	(15.539)
clarity^7	186.742***
	(13.685)
cut.L	713.804***
	(22.511)
cut.Q	−334.503***
	(19.828)
cut.C	188.482***
	(17.218)
cut^4	1.663
	(13.794)
carat	8,472.026***
	(12.615)
Constant	−3,187.540***
	(14.475)
Observations	53,940
R^2	0.897
Adjusted R^2	0.897
Residual Std. Error	1,280.935 (df = 53927)
F Statistic	39,106.570*** (df = 12; 53927)
Note:	*p<0.1; **p<0.05; ***p<0.01