RMarkdown Introduction

Tatjana Kecojević

2018-06-20

R Markdown

This is my first 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.

Material from tonight's workshop can be found on our GitHub page.

When analysing data, a starting point is to examine the characteristics of each individual variable in the data set. The way to proceed depends upon the type of variable being examined. The variables can be one of two broad types:

1. **Attribute variable**: has its outcomes described in terms of its characteristics or attributes;
2. **Measured variable**: has the resulting outcome expressed in numerical terms.

Including R Code

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:

```
install.packages("dplyr", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##  /var/folders/71/96w85flx3yl928r2hzpfwvd00000gp/T//RtmpBv8CsJ/downloaded_packages
```

```
install.packages("ggplot2", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##  /var/folders/71/96w85flx3yl928r2hzpfwvd00000gp/T//RtmpBv8CsJ/downloaded_packages
```

```
install.packages("gapminder", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##  /var/folders/71/96w85flx3yl928r2hzpfwvd00000gp/T//RtmpBv8CsJ/downloaded_packages
```

```
library(gapminder)
summary(gapminder)
```

```
##         country        continent        year         lifeExp
##  Afghanistan:  12   Africa  :624   Min.   :1952   Min.   :23.60
##  Albania    :  12   Americas:300   1st Qu.:1966   1st Qu.:48.20
##  Algeria    :  12   Asia    :396   Median :1980   Median :60.71
##  Angola     :  12   Europe  :360   Mean   :1980   Mean   :59.47
##  Argentina  :  12   Oceania : 24   3rd Qu.:1993   3rd Qu.:70.85
##  Australia  :  12                  Max.   :2007   Max.   :82.60
##  (Other)    :1632
##       pop              gdpPercap
##  Min.   :6.001e+04   Min.   :   241.2
```

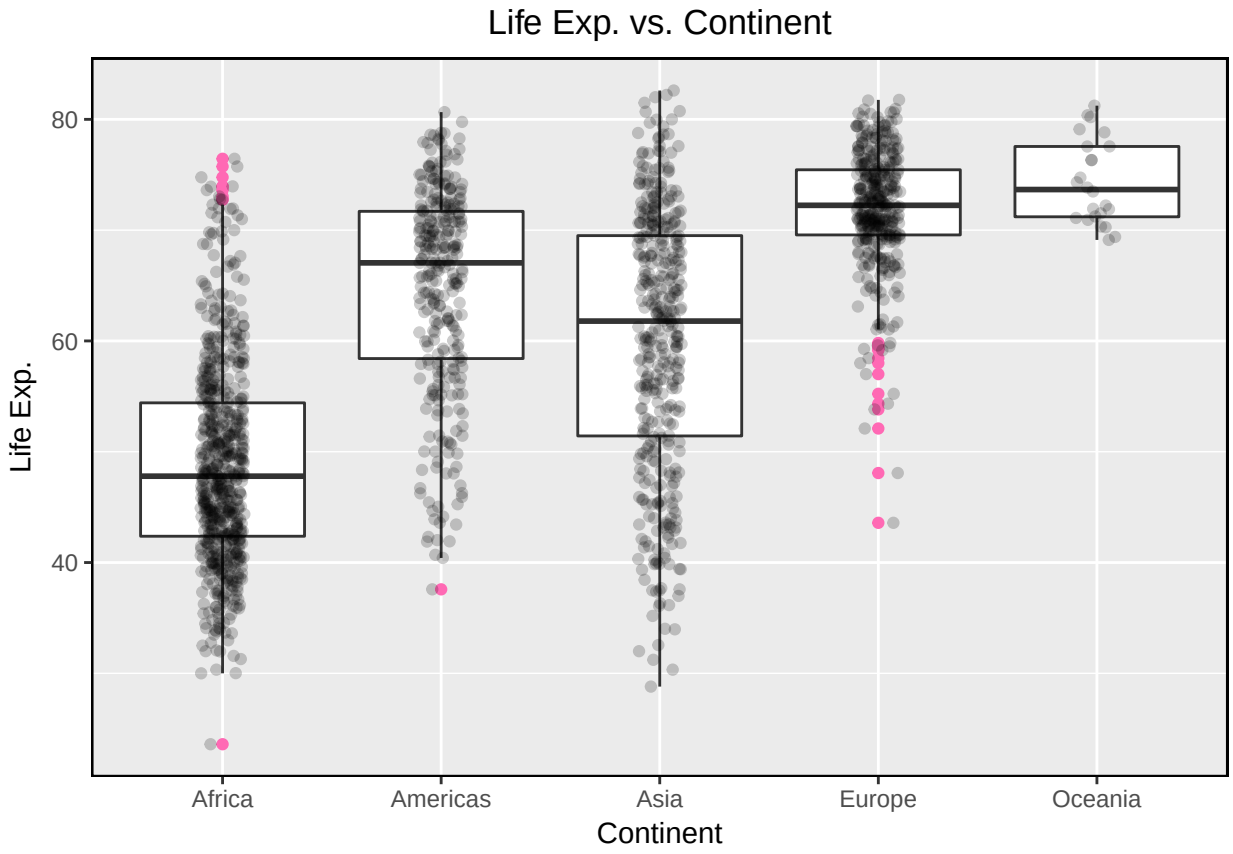
```
##  1st Qu.:2.794e+06   1st Qu.:  1202.1
##  Median :7.024e+06   Median :  3531.8
##  Mean   :2.960e+07   Mean   :  7215.3
##  3rd Qu.:1.959e+07   3rd Qu.:  9325.5
##  Max.   :1.319e+09   Max.   :113523.1
##
```

```
gapminder[1:10, ]
```

```
## # A tibble: 10 x 6
##    country     continent  year lifeExp      pop gdpPercap
##    <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
##  1 Afghanistan Asia       1952    28.8  8425333      779.
##  2 Afghanistan Asia       1957    30.3  9240934      821.
##  3 Afghanistan Asia       1962    32.0 10267083      853.
##  4 Afghanistan Asia       1967    34.0 11537966      836.
##  5 Afghanistan Asia       1972    36.1 13079460      740.
##  6 Afghanistan Asia       1977    38.4 14880372      786.
##  7 Afghanistan Asia       1982    39.9 12881816      978.
##  8 Afghanistan Asia       1987    40.8 13867957      852.
##  9 Afghanistan Asia       1992    41.7 16317921      649.
## 10 Afghanistan Asia       1997    41.8 22227415      635.
```

Including Plots

You can also embed plots by setting `echo = FALSE` to the code chunk to prevent printing of the R code that generates the plot. For example:



Including Mathematical Equations

Let us fit the following model

$$lifeExp = b_0 + b_1 pop + b_2 gdpPercap$$

which we write using the LaTeX.

```
m1 <- lm(gapminder$lifeExp ~ gapminder$pop + gapminder$gdpPercap)
summary(m1)
```

```
##
## Call:
## lm(formula = gapminder$lifeExp ~ gapminder$pop + gapminder$gdpPercap)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -82.754  -7.745   2.055   8.212  18.534
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         5.365e+01  3.225e-01  166.36  < 2e-16 ***
## gapminder$pop       9.728e-09  2.385e-09    4.08 4.72e-05 ***
## gapminder$gdpPercap 7.676e-04  2.568e-05   29.89  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.44 on 1701 degrees of freedom
## Multiple R-squared:  0.3471, Adjusted R-squared:  0.3463
## F-statistic: 452.2 on 2 and 1701 DF,  p-value: < 2.2e-16
```

What do we think of this model?

Let's discuss it next time we meet up.