Introduction to R

Erik van Tilburg Bernardes - ArrietaLab - University of Calgary

3-5 April 2023

## 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:

LETS START!!!

# Installation and loading of the libraries  
# install.packages("rmarkdown")  
library(rmarkdown)

Objects on R

#Create object "x"  
x <- 5  
x

## [1] 5

#you just assigned value "5" to an object called "x"

#create additional objects  
#objects can be made up of alphanumeric characters but can't start with a number.   
#Object names are case sensitive.  
  
y <- 2  
students <- 20  
A <- "Fiocruz"  
y

## [1] 2

#print the values for the different objects created

Operations in R

#operations  
#There are numerous operators in R - arithmetic, logical tests, etc.  
# you have created objects x and y before  
# now create an object that is the sum of them  
z <- x + y  
z

## [1] 7

#or you can just use them directly for different operations   
x \* y

## [1] 10

# value if you multiply them

# let's use some logic now  
x > y

## [1] TRUE

#what about x and z?

Functions

# let's run function "ls()"  
ls()

## [1] "A" "students" "x" "y" "z"

#what this function does is to list all the objects and functions created in your environment

#let's explore the "help()" function  
# on it's own it displays documentation for itself...try it...  
help()

#you can also add the name of another function to explore what it does  
#let's get more details on the "ls()" function we just used  
help(ls) # same as '?ls'

#let's explore other built-in functions  
mean(x, y, z)

## [1] 5

min(x, y, z)

## [1] 2

print(x + y + z)

## [1] 14

sum(x, y, z)

## [1] 14

#let's create our own function  
MySequence <- seq(1, 100)  
MySequence #print your new created sequence

## [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18  
## [19] 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36  
## [37] 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54  
## [55] 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72  
## [73] 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90  
## [91] 91 92 93 94 95 96 97 98 99 100

MyMean <- mean(MySequence)  
MyMean

## [1] 50.5

SumMySequence <- sum(MySequence)  
SumMySequence

## [1] 5050

print(sum(MySequence))

## [1] 5050

vector objects

# a vector is the simplest type of object in R.  
# it is an object that contains one or more values of the same type. Some common types of vector are:  
  
vec.a <- c(x, y, z)  
vec.a

## [1] 5 2 7

#so now creating a new vector  
vec.b <- c("x", "y", "z")  
vec.b #is it the same? what is different?

## [1] "x" "y" "z"

Matrix objects

#A matrix is a two-dimensional numerical array  
#let's create a matrix  
mat.a <- matrix(c(1, 2, 3, 4, 5, 6), nrow = 2, ncol = 3)  
mat.a

## [,1] [,2] [,3]  
## [1,] 1 3 5  
## [2,] 2 4 6

#let's name the rows and columns of the new matrix  
rownames(mat.a) <- c("row1", "row2")  
mat.a

## [,1] [,2] [,3]  
## row1 1 3 5  
## row2 2 4 6

colnames(mat.a) <- c("col1", "col2", "col3")  
mat.a

## col1 col2 col3  
## row1 1 3 5  
## row2 2 4 6

data.frame objects

#its an objects, similar to matrix, but its columns can be a mixture of different types of data.   
# The data.frame is perhaps the most frequently used type of object for biodiversity analysis since it can contain many different types of data.  
  
df.a <- as.data.frame(mat.a)  
df.a$col4 <- c("foo", "bar")  
dim(df.a)

## [1] 2 4

## [1] 2 4  
df.a

## col1 col2 col3 col4  
## row1 1 3 5 foo  
## row2 2 4 6 bar

List objects

#A list is a collection of different types of objects.   
# make a list combining the vector, matrix, and data.frame objects we just created  
list.a <- list(va = vec.a, vb = vec.b, ma = mat.a, dfa = df.a)  
names(list.a)

## [1] "va" "vb" "ma" "dfa"

list.a

## $va  
## [1] 5 2 7  
##   
## $vb  
## [1] "x" "y" "z"  
##   
## $ma  
## col1 col2 col3  
## row1 1 3 5  
## row2 2 4 6  
##   
## $dfa  
## col1 col2 col3 col4  
## row1 1 3 5 foo  
## row2 2 4 6 bar

Tidyverse

#let's install this new package  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.0 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.1 ✔ tibble 3.2.0  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors

#note that tidyverse comes wiht several other packages attached to it

help(tidyverse)

#import global variation in atopy and asthma file  
getwd()

## [1] "/Users/monaparizadeh/Downloads"

atopyasthmavariation <- read.table("WorldwideVariationAtopy.tsv", header=TRUE, sep="\t", dec=".", strip.white=TRUE)  
atopyasthmavariation

## Continent Country Region\_Centre YearoftheStudy  
## 1 Africa Ghana Kintampo 2000  
## 2 Africa Libya Al Aziza 2001  
## 3 Africa Libya Samno 2001  
## 4 Africa Uganda Entebee 2015  
## 5 Africa Ghana Kumasi 2003  
## 6 Africa South Africa Cape Town 2003  
## 7 Africa Kenya Kabati 2002  
## 8 Africa Kenya Thika 2002  
## 9 Africa Nigeria Ojo 1999  
## 10 Asia China San Bu 1992  
## 11 Asia China Beijing 1998  
## 12 Asia China Guangzhou 1998  
## 13 Asia China Hong Kong 1998  
## 14 Asia India Mumbai 2001  
## 15 Asia Malaysia Kota Kinabalu 1992  
## 16 Asia Korea Cheju Island 1998  
## 17 Europe Albania Tirana 1999  
## 18 Europe Denmark Copenhagen 2001  
## 19 Europe Estonia Tallinn 1997  
## 20 Europe Finland Kuopio 2001  
## 21 Europe Georgia Tbilisi 2002  
## 22 Europe Germany Dresden 1996  
## 23 Europe Germany Leipzig, Halle 1992  
## 24 Europe Germany Munich 1996  
## 25 Europe Greece Crete 2001  
## 26 Europe Greece Athens 2001  
## 27 Europe Greece Thessaloniki 2001  
## 28 Europe Iceland Reykjavik 2000  
## 29 Europe Italy Guardea 2001  
## 30 Europe Italy Ronciglione 2003  
## 31 Europe Italy Rome 2001  
## 32 Europe Latvia Riga 1999  
## 33 Europe Norway Oslo 2002  
## 34 Europe Norway Tromso 2000  
## 35 Europe Poland Legnica 2001  
## 36 Europe Poland Starachowice 2001  
## 37 Europe Spain Almeria 2001  
## 38 Europe Spain Cartagena 2001  
## 39 Europe Spain Madrid 2002  
## 40 Europe Spain Valencia 2001  
## 41 Europe Sweden Norrbotten 1996  
## 42 Europe Sweden Ostersund 1997  
## 43 Europe Sweden Umea 1987  
## 44 Europe Sweden Linkoping 1997  
## 45 Europe Sweden Ostersund 1997  
## 46 Europe The Netherlands The Netherlands 1998  
## 47 Europe United Kingdom Ashford 1993  
## 48 Europe United Kingdom Bristol 1999  
## 49 Europe United Kingdom Isle of Wight 1999  
## 50 Europe United Kingdom West Sussex 1999  
## 51 Middle East Turkey Ankara 2000  
## 52 Middle East West Bank Ramallah 2000  
## 53 Middle East Turkey Afyon 2001  
## 54 North America United States Boston 1996  
## 55 North America United States Detroit 1989  
## 56 North America United States Tucson 1984  
## 57 North America Canada Quebec City 2000  
## 58 Oceania Australia Belmont 1984  
## 59 Oceania Australia Canberra 1999  
## 60 Oceania Australia Perth 2004  
## 61 Oceania Australia Villawood 1986  
## 62 Oceania Australia Wagga Wagga 1983  
## 63 Oceania Fiji Suva City 1990  
## 64 Oceania New Zealand Dannevirke 2002  
## 65 Oceania New Zealand Dunedin 1985  
## 66 Oceania New Zealand Hawkes Bay 2000  
## 67 Oceania New Zealand Hastings, Havelock North 2000  
## 68 Oceania Australia Sidney 1999  
## 69 South America Brazil Salvador 2005  
## 70 South America Brazil Uruguaiana 2004  
## 71 South America Ecuador Pichincha & Esmeraldas 2003  
## 72 South America Peru Lima 1997  
## 73 South America Cuba San Juan y Martinez 2004  
## 74 North America United States South Dakota 1997  
## CharacteristicoftheStudyArea Age\_Group NumParticipants Asthma\_Ever Atopy  
## 1 Rural 8-12 1354 15.80 1.7  
## 2 Urban/Rural 7-11 154 9.10 5.8  
## 3 Urban/Rural 7-11 180 4.40 6.7  
## 4 Urban/Rural 0-9 2345 0.80 25.0  
## 5 Urban/Rural 9-16 1848 5.20 13.6  
## 6 Urban 6-14 359 13.60 18.4  
## 7 Rural 9-15 136 3.90 10.9  
## 8 Urban 8-13 129 10.20 25.2  
## 9 Urban/Rural 8-11 566 6.00 28.2  
## 10 Urban/Rural 12-18 737 1.60 49.0  
## 11 Urban 8-12 4214 6.40 23.9  
## 12 Urban 8-12 3510 4.40 32.0  
## 13 Urban 8-12 3011 7.90 45.3  
## 14 Urban 8-12 1658 4.80 6.4  
## 15 Urban/Rural 12-18 409 3.30 63.9  
## 16 Rural 16-18 2005 2.40 36.3  
## 17 Urban 8-12 1052 2.70 15.0  
## 18 Urban 7-17 480 11.90 19.4  
## 19 Urban 8-12 971 2.50 14.6  
## 20 Rural 6-13 765 6.10 33.9  
## 21 Urban 8-12 1012 3.20 33.0  
## 22 Urban 8-12 3023 3.60 25.7  
## 23 Urban 9-10 3105 7.20 18.2  
## 24 Urban 8-12 3301 4.80 22.3  
## 25 Rural 7-18 797 4.40 24.0  
## 26 Urban 8-12 985 7.50 14.4  
## 27 Urban 8-12 1018 11.60 26.8  
## 28 Urban 8-12 937 22.90 23.5  
## 29 Urban/Rural 9-11 101 26.70 31.7  
## 30 Urban/Rural 8-11 166 18.90 31.7  
## 31 Urban 8-12 1354 14.30 28.9  
## 32 Urban 8-12 908 3.20 19.3  
## 33 Urban 10 3754 20.20 29.3  
## 34 Urban/Rural 8-12 3669 10.30 32.7  
## 35 Urban/Rural 8-12 150 27.30 18.0  
## 36 Urban/Rural 9-10 112 28.60 17.9  
## 37 Urban 8-12 1126 14.60 43.0  
## 38 Urban 8-12 1429 10.90 23.8  
## 39 Urban 8-12 981 11.40 34.5  
## 40 Urban 8-12 1362 9.80 14.3  
## 41 Urban/Rural 7-8 3525 8.00 20.6  
## 42 Urban 10-11 1197 10.90 26.8  
## 43 Urban 14 1159 11.00 43.2  
## 44 Urban 8-12 907 9.60 19.8  
## 45 Urban/Rural 8-12 1195 10.90 26.5  
## 46 Urban 8-12 3541 7.80 30.9  
## 47 Urban 0-5 625 17.40 16.7  
## 48 Urban/Rural 0-7 13971 20.00 20.6  
## 49 Urban/Rural 1-10 1456 21.40 29.9  
## 50 Urban/Rural 8-12 1056 20.30 17.5  
## 51 Urban 8-12 3041 6.90 20.6  
## 52 Urban/Rural 8-12 2304 9.40 10.3  
## 53 Urban 13-18 1366 7.46 15.7  
## 54 Urban 0-7 498 11.10 55.0  
## 55 Urban 6-7 825 10.50 33.6  
## 56 Urban 0-6 1246 29.70 41.3  
## 57 Rural 12-19 9082 8.10 44.8  
## 58 Urban 8-10 993 15.80 29.3  
## 59 Urban/Rural 9-10 935 34.30 45.6  
## 60 Urban 0-11 253 14.70 51.4  
## 61 Urban 8-11 1217 15.30 32.8  
## 62 Urban 8-10 1371 19.60 30.6  
## 63 Urban 9-10 2117 8.40 40.9  
## 64 Urban/Rural 7-10 605 33.80 32.4  
## 65 Urban 3-13 1661 44.10 44.8  
## 66 Urban/Rural 8-12 1320 35.60 34.5  
## 67 Urban 11-12 1321 35.70 34.7  
## 68 Urban/Rural 8-11 654 28.00 32.5  
## 69 Urban 4-11 1168 23.30 38.1  
## 70 Urban 8-12 1971 12.70 13.3  
## 71 Rural 5-18 4431 10.30 18.2  
## 72 Urban 8-10 793 20.70 23.9  
## 73 Urban/Rural 5-14 1011 32.70 21.5  
## 74 Rural 6-20 269 9.30 54.3

# read.table is a function to read a data table in 'tsv' format that we have in our workspace  
# part of the function is to tell if the table has a header (=TRUE)  
# sep indicated field separator between characters  
# dec specifies the decimal points used in the file (it might be using ",")  
# strip.white required as we specified "sep" that will account for blank spaces between characters

# Let's explore the new imported data.frame  
colnames(atopyasthmavariation)

## [1] "Continent" "Country"   
## [3] "Region\_Centre" "YearoftheStudy"   
## [5] "CharacteristicoftheStudyArea" "Age\_Group"   
## [7] "NumParticipants" "Asthma\_Ever"   
## [9] "Atopy"

dim(atopyasthmavariation)

## [1] 74 9

# too much information, what can we do?  
#let's filter and sort the data:  
  
AAvariationEurope <- filter(atopyasthmavariation,   
 Continent == "Europe")  
AAvariationEurope

## Continent Country Region\_Centre YearoftheStudy  
## 1 Europe Albania Tirana 1999  
## 2 Europe Denmark Copenhagen 2001  
## 3 Europe Estonia Tallinn 1997  
## 4 Europe Finland Kuopio 2001  
## 5 Europe Georgia Tbilisi 2002  
## 6 Europe Germany Dresden 1996  
## 7 Europe Germany Leipzig, Halle 1992  
## 8 Europe Germany Munich 1996  
## 9 Europe Greece Crete 2001  
## 10 Europe Greece Athens 2001  
## 11 Europe Greece Thessaloniki 2001  
## 12 Europe Iceland Reykjavik 2000  
## 13 Europe Italy Guardea 2001  
## 14 Europe Italy Ronciglione 2003  
## 15 Europe Italy Rome 2001  
## 16 Europe Latvia Riga 1999  
## 17 Europe Norway Oslo 2002  
## 18 Europe Norway Tromso 2000  
## 19 Europe Poland Legnica 2001  
## 20 Europe Poland Starachowice 2001  
## 21 Europe Spain Almeria 2001  
## 22 Europe Spain Cartagena 2001  
## 23 Europe Spain Madrid 2002  
## 24 Europe Spain Valencia 2001  
## 25 Europe Sweden Norrbotten 1996  
## 26 Europe Sweden Ostersund 1997  
## 27 Europe Sweden Umea 1987  
## 28 Europe Sweden Linkoping 1997  
## 29 Europe Sweden Ostersund 1997  
## 30 Europe The Netherlands The Netherlands 1998  
## 31 Europe United Kingdom Ashford 1993  
## 32 Europe United Kingdom Bristol 1999  
## 33 Europe United Kingdom Isle of Wight 1999  
## 34 Europe United Kingdom West Sussex 1999  
## CharacteristicoftheStudyArea Age\_Group NumParticipants Asthma\_Ever Atopy  
## 1 Urban 8-12 1052 2.7 15.0  
## 2 Urban 7-17 480 11.9 19.4  
## 3 Urban 8-12 971 2.5 14.6  
## 4 Rural 6-13 765 6.1 33.9  
## 5 Urban 8-12 1012 3.2 33.0  
## 6 Urban 8-12 3023 3.6 25.7  
## 7 Urban 9-10 3105 7.2 18.2  
## 8 Urban 8-12 3301 4.8 22.3  
## 9 Rural 7-18 797 4.4 24.0  
## 10 Urban 8-12 985 7.5 14.4  
## 11 Urban 8-12 1018 11.6 26.8  
## 12 Urban 8-12 937 22.9 23.5  
## 13 Urban/Rural 9-11 101 26.7 31.7  
## 14 Urban/Rural 8-11 166 18.9 31.7  
## 15 Urban 8-12 1354 14.3 28.9  
## 16 Urban 8-12 908 3.2 19.3  
## 17 Urban 10 3754 20.2 29.3  
## 18 Urban/Rural 8-12 3669 10.3 32.7  
## 19 Urban/Rural 8-12 150 27.3 18.0  
## 20 Urban/Rural 9-10 112 28.6 17.9  
## 21 Urban 8-12 1126 14.6 43.0  
## 22 Urban 8-12 1429 10.9 23.8  
## 23 Urban 8-12 981 11.4 34.5  
## 24 Urban 8-12 1362 9.8 14.3  
## 25 Urban/Rural 7-8 3525 8.0 20.6  
## 26 Urban 10-11 1197 10.9 26.8  
## 27 Urban 14 1159 11.0 43.2  
## 28 Urban 8-12 907 9.6 19.8  
## 29 Urban/Rural 8-12 1195 10.9 26.5  
## 30 Urban 8-12 3541 7.8 30.9  
## 31 Urban 0-5 625 17.4 16.7  
## 32 Urban/Rural 0-7 13971 20.0 20.6  
## 33 Urban/Rural 1-10 1456 21.4 29.9  
## 34 Urban/Rural 8-12 1056 20.3 17.5

dim(AAvariationEurope)

## [1] 34 9

# you used the equal to sign "==" to select only the data you wanted

# what is you want to do the oposite? instead of selecting just one, you want to exclude a continent  
AAvariationNoOceania <- filter(atopyasthmavariation,   
 Continent != "Oceania")  
AAvariationNoOceania

## Continent Country Region\_Centre YearoftheStudy  
## 1 Africa Ghana Kintampo 2000  
## 2 Africa Libya Al Aziza 2001  
## 3 Africa Libya Samno 2001  
## 4 Africa Uganda Entebee 2015  
## 5 Africa Ghana Kumasi 2003  
## 6 Africa South Africa Cape Town 2003  
## 7 Africa Kenya Kabati 2002  
## 8 Africa Kenya Thika 2002  
## 9 Africa Nigeria Ojo 1999  
## 10 Asia China San Bu 1992  
## 11 Asia China Beijing 1998  
## 12 Asia China Guangzhou 1998  
## 13 Asia China Hong Kong 1998  
## 14 Asia India Mumbai 2001  
## 15 Asia Malaysia Kota Kinabalu 1992  
## 16 Asia Korea Cheju Island 1998  
## 17 Europe Albania Tirana 1999  
## 18 Europe Denmark Copenhagen 2001  
## 19 Europe Estonia Tallinn 1997  
## 20 Europe Finland Kuopio 2001  
## 21 Europe Georgia Tbilisi 2002  
## 22 Europe Germany Dresden 1996  
## 23 Europe Germany Leipzig, Halle 1992  
## 24 Europe Germany Munich 1996  
## 25 Europe Greece Crete 2001  
## 26 Europe Greece Athens 2001  
## 27 Europe Greece Thessaloniki 2001  
## 28 Europe Iceland Reykjavik 2000  
## 29 Europe Italy Guardea 2001  
## 30 Europe Italy Ronciglione 2003  
## 31 Europe Italy Rome 2001  
## 32 Europe Latvia Riga 1999  
## 33 Europe Norway Oslo 2002  
## 34 Europe Norway Tromso 2000  
## 35 Europe Poland Legnica 2001  
## 36 Europe Poland Starachowice 2001  
## 37 Europe Spain Almeria 2001  
## 38 Europe Spain Cartagena 2001  
## 39 Europe Spain Madrid 2002  
## 40 Europe Spain Valencia 2001  
## 41 Europe Sweden Norrbotten 1996  
## 42 Europe Sweden Ostersund 1997  
## 43 Europe Sweden Umea 1987  
## 44 Europe Sweden Linkoping 1997  
## 45 Europe Sweden Ostersund 1997  
## 46 Europe The Netherlands The Netherlands 1998  
## 47 Europe United Kingdom Ashford 1993  
## 48 Europe United Kingdom Bristol 1999  
## 49 Europe United Kingdom Isle of Wight 1999  
## 50 Europe United Kingdom West Sussex 1999  
## 51 Middle East Turkey Ankara 2000  
## 52 Middle East West Bank Ramallah 2000  
## 53 Middle East Turkey Afyon 2001  
## 54 North America United States Boston 1996  
## 55 North America United States Detroit 1989  
## 56 North America United States Tucson 1984  
## 57 North America Canada Quebec City 2000  
## 58 South America Brazil Salvador 2005  
## 59 South America Brazil Uruguaiana 2004  
## 60 South America Ecuador Pichincha & Esmeraldas 2003  
## 61 South America Peru Lima 1997  
## 62 South America Cuba San Juan y Martinez 2004  
## 63 North America United States South Dakota 1997  
## CharacteristicoftheStudyArea Age\_Group NumParticipants Asthma\_Ever Atopy  
## 1 Rural 8-12 1354 15.80 1.7  
## 2 Urban/Rural 7-11 154 9.10 5.8  
## 3 Urban/Rural 7-11 180 4.40 6.7  
## 4 Urban/Rural 0-9 2345 0.80 25.0  
## 5 Urban/Rural 9-16 1848 5.20 13.6  
## 6 Urban 6-14 359 13.60 18.4  
## 7 Rural 9-15 136 3.90 10.9  
## 8 Urban 8-13 129 10.20 25.2  
## 9 Urban/Rural 8-11 566 6.00 28.2  
## 10 Urban/Rural 12-18 737 1.60 49.0  
## 11 Urban 8-12 4214 6.40 23.9  
## 12 Urban 8-12 3510 4.40 32.0  
## 13 Urban 8-12 3011 7.90 45.3  
## 14 Urban 8-12 1658 4.80 6.4  
## 15 Urban/Rural 12-18 409 3.30 63.9  
## 16 Rural 16-18 2005 2.40 36.3  
## 17 Urban 8-12 1052 2.70 15.0  
## 18 Urban 7-17 480 11.90 19.4  
## 19 Urban 8-12 971 2.50 14.6  
## 20 Rural 6-13 765 6.10 33.9  
## 21 Urban 8-12 1012 3.20 33.0  
## 22 Urban 8-12 3023 3.60 25.7  
## 23 Urban 9-10 3105 7.20 18.2  
## 24 Urban 8-12 3301 4.80 22.3  
## 25 Rural 7-18 797 4.40 24.0  
## 26 Urban 8-12 985 7.50 14.4  
## 27 Urban 8-12 1018 11.60 26.8  
## 28 Urban 8-12 937 22.90 23.5  
## 29 Urban/Rural 9-11 101 26.70 31.7  
## 30 Urban/Rural 8-11 166 18.90 31.7  
## 31 Urban 8-12 1354 14.30 28.9  
## 32 Urban 8-12 908 3.20 19.3  
## 33 Urban 10 3754 20.20 29.3  
## 34 Urban/Rural 8-12 3669 10.30 32.7  
## 35 Urban/Rural 8-12 150 27.30 18.0  
## 36 Urban/Rural 9-10 112 28.60 17.9  
## 37 Urban 8-12 1126 14.60 43.0  
## 38 Urban 8-12 1429 10.90 23.8  
## 39 Urban 8-12 981 11.40 34.5  
## 40 Urban 8-12 1362 9.80 14.3  
## 41 Urban/Rural 7-8 3525 8.00 20.6  
## 42 Urban 10-11 1197 10.90 26.8  
## 43 Urban 14 1159 11.00 43.2  
## 44 Urban 8-12 907 9.60 19.8  
## 45 Urban/Rural 8-12 1195 10.90 26.5  
## 46 Urban 8-12 3541 7.80 30.9  
## 47 Urban 0-5 625 17.40 16.7  
## 48 Urban/Rural 0-7 13971 20.00 20.6  
## 49 Urban/Rural 1-10 1456 21.40 29.9  
## 50 Urban/Rural 8-12 1056 20.30 17.5  
## 51 Urban 8-12 3041 6.90 20.6  
## 52 Urban/Rural 8-12 2304 9.40 10.3  
## 53 Urban 13-18 1366 7.46 15.7  
## 54 Urban 0-7 498 11.10 55.0  
## 55 Urban 6-7 825 10.50 33.6  
## 56 Urban 0-6 1246 29.70 41.3  
## 57 Rural 12-19 9082 8.10 44.8  
## 58 Urban 4-11 1168 23.30 38.1  
## 59 Urban 8-12 1971 12.70 13.3  
## 60 Rural 5-18 4431 10.30 18.2  
## 61 Urban 8-10 793 20.70 23.9  
## 62 Urban/Rural 5-14 1011 32.70 21.5  
## 63 Rural 6-20 269 9.30 54.3

dim(AAvariationNoOceania)

## [1] 63 9

# you used the differ to sign "!=" to select all data but the one you dont want

#Now we want to see the countries with the highest atopy incidence  
HighAtopy <- filter(atopyasthmavariation,   
 Atopy > 25) %>%   
 arrange(Country)  
HighAtopy

## Continent Country Region\_Centre YearoftheStudy  
## 1 Oceania Australia Belmont 1984  
## 2 Oceania Australia Canberra 1999  
## 3 Oceania Australia Perth 2004  
## 4 Oceania Australia Villawood 1986  
## 5 Oceania Australia Wagga Wagga 1983  
## 6 Oceania Australia Sidney 1999  
## 7 South America Brazil Salvador 2005  
## 8 North America Canada Quebec City 2000  
## 9 Asia China San Bu 1992  
## 10 Asia China Guangzhou 1998  
## 11 Asia China Hong Kong 1998  
## 12 Oceania Fiji Suva City 1990  
## 13 Europe Finland Kuopio 2001  
## 14 Europe Georgia Tbilisi 2002  
## 15 Europe Germany Dresden 1996  
## 16 Europe Greece Thessaloniki 2001  
## 17 Europe Italy Guardea 2001  
## 18 Europe Italy Ronciglione 2003  
## 19 Europe Italy Rome 2001  
## 20 Africa Kenya Thika 2002  
## 21 Asia Korea Cheju Island 1998  
## 22 Asia Malaysia Kota Kinabalu 1992  
## 23 Oceania New Zealand Dannevirke 2002  
## 24 Oceania New Zealand Dunedin 1985  
## 25 Oceania New Zealand Hawkes Bay 2000  
## 26 Oceania New Zealand Hastings, Havelock North 2000  
## 27 Africa Nigeria Ojo 1999  
## 28 Europe Norway Oslo 2002  
## 29 Europe Norway Tromso 2000  
## 30 Europe Spain Almeria 2001  
## 31 Europe Spain Madrid 2002  
## 32 Europe Sweden Ostersund 1997  
## 33 Europe Sweden Umea 1987  
## 34 Europe Sweden Ostersund 1997  
## 35 Europe The Netherlands The Netherlands 1998  
## 36 Europe United Kingdom Isle of Wight 1999  
## 37 North America United States Boston 1996  
## 38 North America United States Detroit 1989  
## 39 North America United States Tucson 1984  
## 40 North America United States South Dakota 1997  
## CharacteristicoftheStudyArea Age\_Group NumParticipants Asthma\_Ever Atopy  
## 1 Urban 8-10 993 15.8 29.3  
## 2 Urban/Rural 9-10 935 34.3 45.6  
## 3 Urban 0-11 253 14.7 51.4  
## 4 Urban 8-11 1217 15.3 32.8  
## 5 Urban 8-10 1371 19.6 30.6  
## 6 Urban/Rural 8-11 654 28.0 32.5  
## 7 Urban 4-11 1168 23.3 38.1  
## 8 Rural 12-19 9082 8.1 44.8  
## 9 Urban/Rural 12-18 737 1.6 49.0  
## 10 Urban 8-12 3510 4.4 32.0  
## 11 Urban 8-12 3011 7.9 45.3  
## 12 Urban 9-10 2117 8.4 40.9  
## 13 Rural 6-13 765 6.1 33.9  
## 14 Urban 8-12 1012 3.2 33.0  
## 15 Urban 8-12 3023 3.6 25.7  
## 16 Urban 8-12 1018 11.6 26.8  
## 17 Urban/Rural 9-11 101 26.7 31.7  
## 18 Urban/Rural 8-11 166 18.9 31.7  
## 19 Urban 8-12 1354 14.3 28.9  
## 20 Urban 8-13 129 10.2 25.2  
## 21 Rural 16-18 2005 2.4 36.3  
## 22 Urban/Rural 12-18 409 3.3 63.9  
## 23 Urban/Rural 7-10 605 33.8 32.4  
## 24 Urban 3-13 1661 44.1 44.8  
## 25 Urban/Rural 8-12 1320 35.6 34.5  
## 26 Urban 11-12 1321 35.7 34.7  
## 27 Urban/Rural 8-11 566 6.0 28.2  
## 28 Urban 10 3754 20.2 29.3  
## 29 Urban/Rural 8-12 3669 10.3 32.7  
## 30 Urban 8-12 1126 14.6 43.0  
## 31 Urban 8-12 981 11.4 34.5  
## 32 Urban 10-11 1197 10.9 26.8  
## 33 Urban 14 1159 11.0 43.2  
## 34 Urban/Rural 8-12 1195 10.9 26.5  
## 35 Urban 8-12 3541 7.8 30.9  
## 36 Urban/Rural 1-10 1456 21.4 29.9  
## 37 Urban 0-7 498 11.1 55.0  
## 38 Urban 6-7 825 10.5 33.6  
## 39 Urban 0-6 1246 29.7 41.3  
## 40 Rural 6-20 269 9.3 54.3

# you now selected only the countries with Atopy values higher than 25% of the sample population  
# here we introduce a new language: the pipe operator in the form %>% aims to combine various functions without the need to assign the result to a new object. The pipe operator passes the output of a function applied to the first argument of the next function.   
# using the pipe operator, we applied the filter and the arrange function to the same object HighAtopy

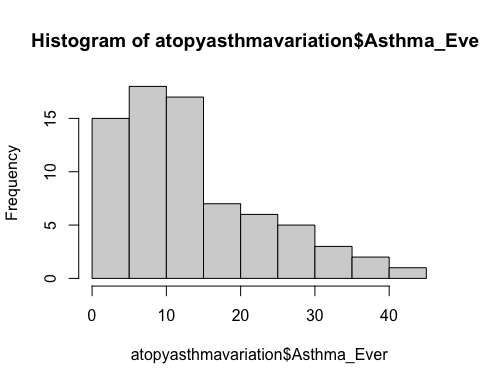
#we can also group informations   
AApercontinent <- atopyasthmavariation %>%   
 group\_by(Continent) %>%   
 summarise(mx = max(Asthma\_Ever, na.rm = TRUE),   
 min = min(Asthma\_Ever, na.rm = TRUE),  
 median = median(Asthma\_Ever))  
  
AApercontinent

## # A tibble: 7 × 4  
## Continent mx min median  
## <chr> <dbl> <dbl> <dbl>  
## 1 Africa 15.8 0.8 6   
## 2 Asia 7.9 1.6 4.4   
## 3 Europe 28.6 2.5 10.9   
## 4 Middle East 9.4 6.9 7.46  
## 5 North America 29.7 8.1 10.5   
## 6 Oceania 44.1 8.4 28   
## 7 South America 32.7 10.3 20.7

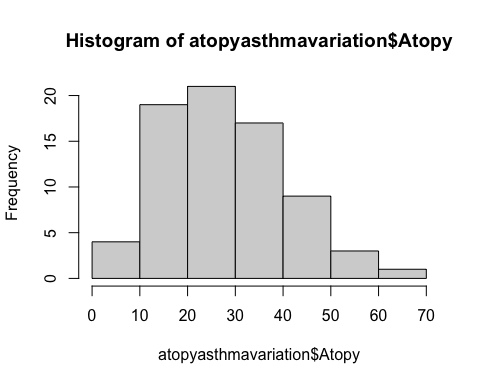
# now we created an object AApercontinent, in which we grouped the Asthma\_Ever incidence per continent  
# in addition, we summarized the data for max, minimum, and median values.

Now lets see the data!

#lets start with some built-in tools  
hist(atopyasthmavariation$Asthma\_Ever)

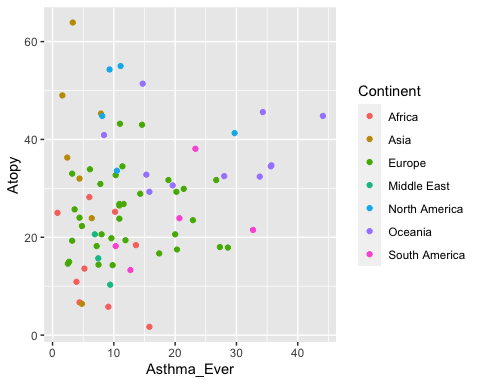


#what about atopy?  
hist(atopyasthmavariation$Atopy)



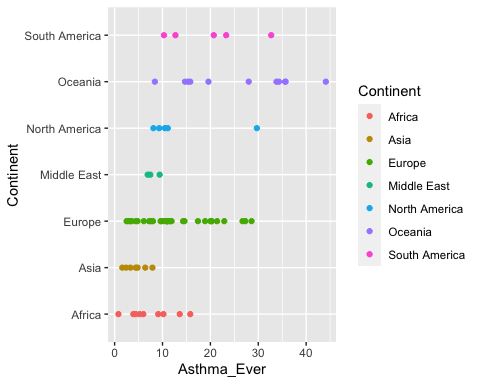
#now let's do more...  
#first install ggplot2  
library(ggplot2)  
#check if it was properly installed  
#also inspect ggplot2 to get more information about the package

plot1 <- ggplot(atopyasthmavariation, aes(Asthma\_Ever, Atopy, colour = Continent)) +   
 geom\_point()  
plot1



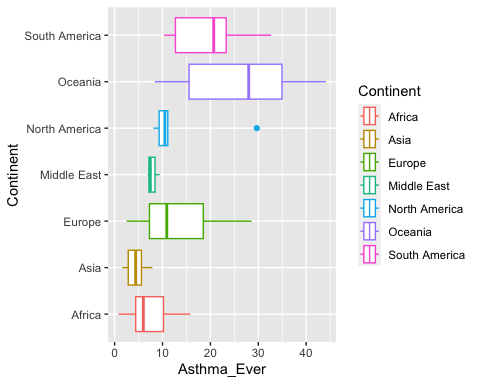
# here we are just plotting our data  
# asthma\_ever vs. atopy, and color coded per continent  
# but as is we can't make much sense of it

plot2 <- ggplot(atopyasthmavariation, aes(Asthma\_Ever, Continent, colour = Continent)) +   
 geom\_point()  
plot2



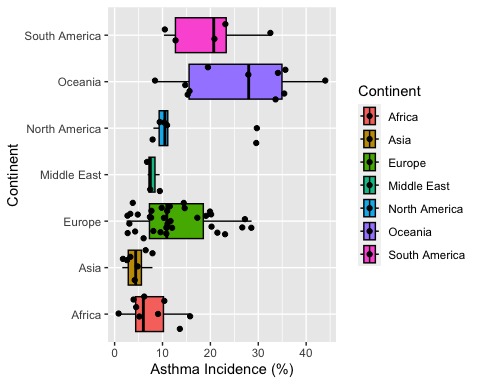
# were we are now looking at asthma incidence per continent

plot3 <- ggplot(atopyasthmavariation, aes(Asthma\_Ever, Continent, colour = Continent)) +   
 geom\_boxplot()  
plot3

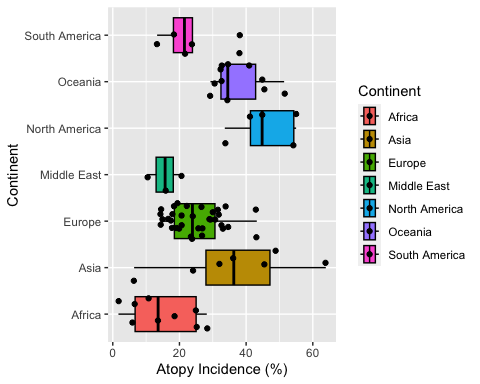


# here we have the same graph...but instead of geom\_points, now we are using geom\_boxplot  
# asthma\_ever vs. atopy, and color coded per continent

# but we mentioned with ggplot we can edit the graphs so they look however we want, we just keeping adding layers to it...so let's explore  
plot4 <- ggplot(atopyasthmavariation, aes(Asthma\_Ever, Continent, fill = Continent, colour = Continent)) + #adding fill, now we also colored the boxplots instead of just outlines  
 geom\_boxplot(alpha=1) + #keeping the assigned boxplots for the graph  
 geom\_jitter(width=0.2) + #but we also want to see the individual points, not just the bars  
 xlab("Asthma Incidence (%)") + #now we are renaming the x axis  
 scale\_color\_manual(name="Continent", values=c("black", "black", "black", "black", "black", "black", "black"))   
plot4



# now let's do the same for atopy  
plot5 <- ggplot(atopyasthmavariation, aes(Atopy, Continent, fill = Continent, colour = Continent)) +   
 geom\_boxplot(alpha=1) +  
 geom\_jitter(width=0.2) +   
 xlab("Atopy Incidence (%)") + #here we need to manually change the axis name  
 scale\_color\_manual(name="Continent", values=c("black", "black", "black", "black", "black", "black", "black"))   
plot5



# and another nice thing we can do is to combine graphs to generate a report  
# we need to install a new package though  
library(gridExtra)

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

grid.arrange(nrow = 2, plot4, plot5)

