

Introduction

R's basic data types are character, numeric, integer, complex, and logical.

R's basic data structures include the vector, list, matrix, data frame, and factors. Some of these structures require that all members be of the same data type (e.g. vectors, matrices) while others permit multiple data types (e.g. lists, data frames)

Data types

Variables in R can be of different types. For example, we need to distinguish numbers from character strings and tables from simple lists of numbers. The function `class` helps us determine what type of object we have:

```
a <- 2
class(a)
```

```
## [1] "numeric"
```

To work efficiently in R it is important to learn the different types of variables and what we can do with these.

Data Frames

Up to now, the variables we have defined are just one number. This is not very useful for storing data. The most common way of storing a dataset in R is in a *data frame*. Conceptually, we can think of a data frame as a table with rows representing observations and the different variables reported for each observation defining the columns. Data frames are particularly useful for datasets because we can combine different data types into one object.

We stored the data for our motivating example in a data frame. You can access this dataset by loading the `dslabs` library and loading the `murders` dataset using the `data` function:

```
library(dslabs)
```

```
## Warning: package 'dslabs' was built under R version 4.1.3
```

```
data(murders)
```

To see that this is in fact a data frame we type

```
class(murders)
```

```
## [1] "data.frame"
```

Examining an object

The function `str` is useful to find out more about the structure of an object

```
str(murders)
```

```
## 'data.frame':   51 obs. of  5 variables:
## $ state      : chr  "Alabama" "Alaska" "Arizona" "Arkansas" ...
## $ abb       : chr  "AL" "AK" "AZ" "AR" ...
## $ region    : Factor w/ 4 levels "Northeast","South",...: 2 4 4 2 4 4 1 2 2 2 ...
## $ population: num  4779736 710231 6392017 2915918 37253956 ...
## $ total     : num  135 19 232 93 1257 ...
```

This tells us much more about the object. We see that the table has 51 rows (50 states plus DC) and five variables. We can show the first six lines using the function `head`:

```
head(murders)
```

```
##      state abb region population total
## 1   Alabama AL  South   4779736    135
## 2    Alaska AK   West    710231     19
## 3   Arizona AZ   West    6392017    232
## 4  Arkansas AR  South    2915918     93
## 5 California CA  West    37253956   1257
## 6   Colorado CO   West    5029196     65
```

Before we go any further in answering our original question about different states, let's get to know the components of this object better.

The accessor

For our analysis we will need to access the different variables, represented by columns, included in this data frame. To access these variables we use the accessor operator `$` in the following way:

```
murders$population
```

```
## [1] 4779736 710231 6392017 2915918 37253956 5029196 3574097 897934
## [9] 601723 19687653 9920000 1360301 1567582 12830632 6483802 3046355
## [17] 2853118 4339367 4533372 1328361 5773552 6547629 9883640 5303925
## [25] 2967297 5988927 989415 1826341 2700551 1316470 8791894 2059179
## [33] 19378102 9535483 672591 11536504 3751351 3831074 12702379 1052567
## [41] 4625364 814180 6346105 25145561 2763885 625741 8001024 6724540
## [49] 1852994 5686986 563626
```

But how did we know to use `population`? Above, by applying the function `str` to the object `murders`, we revealed the names for each of the five variables stored in this table. We can quickly access the variables names using:

```
names(murders)
```

```
## [1] "state"      "abb"        "region"     "population" "total"
```

It is important to know that the order of the entries in `murders$population` preserve the order of the rows in our data table. This will later permit us to manipulate one variable based on the results of another. For example, we will be able to order the state names by number of murders.

Tip: R comes with a very nice auto-complete functionality that saves us the trouble of typing out all the names. Try typing `murders$p` then hitting the `tab` key on your keyboard. RStudio has many useful auto-complete feature options.

Vectors: numerics, characters, and logical

Note that the object `murders$population` is not one number but several. We call these types of objects *vectors*. A single number is technically a vector but in general vectors refer to objects with several entries. The function `length` tells you how many entries are in the vector:

```
pop <- murders$population
length(pop)
```

```
## [1] 51
```

This particular vector is *numeric* since population sizes are numbers:

```
class(pop)
```

```
## [1] "numeric"
```

In a numeric vector, every entry must be a number.

To store character strings, vectors can also be of class *character*. For example, the state names are characters:

```
class(murders$state)
```

```
## [1] "character"
```

Most Tweets posted on twitter are also character strings.

```
## E.g what guys are posting now concerning floods on twitter in South Africa
library(rtweet)
```

```
## Warning: package 'rtweet' was built under R version 4.1.3
```

```
rds <- search_tweets("floods")
rds$text
```

```
## [1] "18 people have died in Assam & 7 lakh have been affected due to floods but he's busy with
## [2] "Floods kill 25 in India's Assam, displace thousands https://t.co/z588m9zhbe"
## [3] "In view of the first wave of flood and destruction it caused to the hill district, we visited
## [4] "No bi small floods people go experience https://t.co/UlhC5KZPk9"
## [5] "Apocalyptic scenes in India and Bangladesh\n&3 million displaced by #floods \n&Over 60 c
## [6] "No bi small floods people go experience https://t.co/UlhC5KZPk9"
## [7] "@Smokey058 @Our_DA The People on the Cape Flats as well as many of those affected by the flood
## [8] "ACCRA FLOODS: Korle lagoon and the Odaw drains to be transformed into a recreational center c
## [9] "Our @OntLiberal plan will invest to protect communities against increasingly intense and frequ
## [10] "<U+0001F4A6> The 3rd Cavalry Battalion, 1st Cavalry Regiment deployed Army volunteers to prev
## [11] "Waiving with a small <U+0001F4E6> <U+0001F447>but an important item inside! As part of effort
## [12] "Accra Floods: The situation here in Tsado is serious, this dawn I had to get out of my house
## [13] "DA calls for urgent update on emergency housing following weekend floods https://t.co/RBtCbpij
## [14] "June rains won't be as bad... no floods\n\nJune rains: https://t.co/hdFlMWNjqg"
## [15] "This is insane -\nI'm sure I heard green bandt on Sky say that coal mining and gas exploratio
## [16] "@Afotey_jr @benbaafi You niggas for no dey drive through floods oo yooo. Then Ablorh dey talk
```

[17] "https://t.co/XsJL30A59v Sir I think this is high time you postpone the #NEETPG2022 exam.This :
[18] "As floods continue to create havoc in the state of Assam, @BJYMAssamPrdsh has been continuous:
[19] "Delighted to share our 1st paper now open access in Transactions @RGS_IBG \n\nIn it, we \"fol.
[20] "About six weeks after floods devastated parts of KwaZulu-Natal and the Eastern Cape, MPs want
[21] "@Hitz1039FM The floods situation in Accra is quite disturbing, we are still in this challenge
[22] "Is there a shelter to help people with the floods?"
[23] "Many parts of Bangladesh and neighbouring regions in India are prone to flooding, and scienti
[24] "BLOCKED ROADS AT TESHIE MOBILE DUE TO FLOODS.\nUSE ALTERNATE ROUTES"
[25] "To whom it may concern? \n\n#Labor Minister for #Housing and #Homelessness\n\nCurrently @Jaso
[26] "We are going to wake up to horrendous stories caused by floods."
[27] "@cinveen @YFKoduahSarpong And that place also floods. Sad. Hope you are safe"
[28] "ACCRA FLOODS: Korle lagoon and the Odaw drains to be transformed into a recreational center c
[29] "Disaster don't Differentiate \nDoctors For You treat everyone equally. Today we distributed F
[30] "We are going to wake up to horrendous stories caused by floods."
[31] "There's a 50% chance we'll hit 1.5C by 2027, & the warming won't stop there, it will get v
[32] "Re: the Durban Floods, it's not entirely climate change. If you allow your city to endlessly s
[33] "About six weeks after floods devastated parts of KwaZulu-Natal and the Eastern Cape, MPs want
[34] "KZN floods have exposed our engineers. Tenderpreneurs are not engineers."
[35] "ACCRA FLOODS: Korle lagoon and the Odaw drains to be transformed into a recreational center c
[36] "18 people have died in Assam & 7 lakh have been affected due to floods but he's busy with
[37] "Many parts of Bangladesh and neighbouring regions in India are prone to flooding, and scienti
[38] "We are going to wake up to horrendous stories caused by floods."
[39] "Four million people hit by floods in Bangladesh: UN https://t.co/9w01Rciqkx"
[40] "Assam CM Himanta Biswa Sarma today visited Haflong, the district headquarters of Dima Hasao &
[41] "https://t.co/T5PRGZWMol"
[42] "@miffyjayne We won't cope with more fires and floods - stupid clod."
[43] "In view of the first wave of flood and destruction it caused to the hill district, we visited
[44] "So far 25 people dead and over 1 million people affected due to flash floods in North East In
[45] "About six weeks after floods devastated parts of KwaZulu-Natal and the Eastern Cape, MPs want
[46] "18 people have died in Assam & 7 lakh have been affected due to floods but he's busy with
[47] "How do we move from 0-1 when all our energy and focus is channeled towards negative 1s and 2s
[48] "There's a 50% chance we'll hit 1.5C by 2027, & the warming won't stop there, it will get v
[49] "We are going to wake up to horrendous stories caused by floods."
[50] "About six weeks after floods devastated parts of KwaZulu-Natal and the Eastern Cape, MPs want
[51] "18 people have died in Assam & 7 lakh have been affected due to floods but he's busy with
[52] "@Fithissey @Mountaingoat55 I believe the people of Lismore were abandoned by the Nationals du
[53] "Unstoppable Floods Of Recurring Income Day After Day\nhttps://t.co/JT76yjLfxE #AffiliateMarke
[54] "Charley as e dey rain from yesterday till now Wey I say make I check what's up plus floods. W
[55] "Assam CM Himanta Biswa Sarma today visited Haflong, the district headquarters of Dima Hasao &
[56] "Troops had relocated the examination center to a new location to avoid the inconvenience caus
[57] "Bangladesh and India hit by the worst floods in decades\n\nhttps://t.co/JQwdU3n3st https://t.
[58] "@Hitz1039FM The floods situation in Accra is quite disturbing, we are still in this challenge
[59] "Accra Floods again & This won't be the end - Everyone's to blame but poor leadership has
[60] "BLOCKED ROADS AT TESHIE MOBILE DUE TO FLOODS.\nUSE ALTERNATE ROUTES"
[61] "ACCRA FLOODS LOOMING ONCE AGAIN<U+0001F926><U+0001F3FB><U+200D><U+2642><U+FE0F>"
[62] "Cyclones to double in climate crisis
[63] "#ClimateChange is primarily a #WaterCrisis. We feel its impacts through worsening floods, ris
[64] "Mop-up operations under way after second wave of floods in KZN https://t.co/CNhJQr3up2"
[65] "We are going to wake up to horrendous stories caused by floods."
[66] "As floods continue to create havoc in the state of Assam, @BJYMAssamPrdsh has been continuous:
[67] "#Facts #FactsFirst\n\nIf We Have Sufficient Trees, We Can Have Sufficient Rains & Our Ri
[68] "Inspected the flooded areas of Manduli in Lanka,Hojai along with MLA Shri @RK GhoshBJP today w
[69] "#Facts #FactsFirst\n\nIf We Have Sufficient Trees, We Can Have Sufficient Rains & Our Ri
[70] "In view of the first wave of flood and destruction it caused to the hill district, we visited

```
## [71] "We are going to wake up to horrendous stories caused by floods."
## [72] "We are going to wake up to horrendous stories caused by floods."
## [73] "Mop-up operations under way after second wave of floods in KZN https://t.co/CNhJQr3up2"
## [74] "Power generation to dwindle as 23% of dams exposed to droughts, 54% to floods over climate ri
## [75] "Over 25 dead in floods in Assam and he's away spouting rubbish https://t.co/KDTnuMf6ST"
## [76] "Aside from the Bengali news, this is the first bit of media coverage I've seen since the flood
## [77] "@Birmo Those still camping out after the fires, and those still without homes after the flood
## [78] "ACCRA FLOODS LOOMING ONCE AGAIN<U+0001F926><U+0001F3FB><U+200D><U+2642><U+FE0F>"
## [79] "We are going to wake up to horrendous stories caused by floods."
## [80] "If we really want to stop floods in Ghana especially Accra , one thing that need to be stoppe
## [81] "BLOCKED ROADS AT TESHIE MOBILE DUE TO FLOODS.\nUSE ALTERNATE ROUTES"
## [82] "We are going to wake up to horrendous stories caused by floods."
## [83] "@mukesh_halari @Dr_Uditraj @INCIndia Kabhi yeh sunayi detaa hai? 2014 Uttarakhand floods main
## [84] "We are going to wake up to horrendous stories caused by floods."
## [85] "We are going to wake up to horrendous stories caused by floods."
## [86] "@thenanaaba naa no floods. it didn't rain heavily like 3 days ago."
## [87] "18 people have died in Assam & 7 lakh have been affected due to floods but he's busy with
## [88] "@archillect Just recently (last year) my drunken *ss of a cousin lied to me that we were going
## [89] "KZN Cogta says 48 people missing from April floods now presumed dead https://t.co/mkCr5zRPqy 1
## [90] "LPG prices, jobs, Assam floods - Varanasi Muslims say Gyanvapi row a 'bogey' to distract http
## [91] "<U+0001F3DD><U+FE0F>As the waters have actually declined after the floods have actually finis
## [92] "A broken sea wall caused flooding in Semarang today, residents began to be evacuated. Central
## [93] "@HinduInfopedia @JaipurDialogues @Newsumindia @SadhguruJV Please studey how fertile agricultu
## [94] "18 people have died in Assam & 7 lakh have been affected due to floods but he's busy with
## [95] "#Facts #FactsFirst\n \nIf We Have Sufficient Trees, We Can Have Sufficient Rains & Our Ri
## [96] "GREY LIVE?! CATCH ME IN FLOODS OF TEARS"
## [97] "More Disastrous Floods. More lower atmospheric energy and more water vapor- more disastrous f
## [98] "We are going to wake up to horrendous stories caused by floods."
## [99] "Assam: 25 Dead as Floods Wreak Havoc, 6.5 Lakh People Affected https://t.co/q3MGQQct7A"
## [100] "BLOCKED ROADS AT TESHIE MOBILE DUE TO FLOODS.\nUSE ALTERNATE ROUTES"
```

As with numeric vectors, all entries in a character vector need to be a character.

Another important type are *logical vectors*. These must be either TRUE or FALSE.

```
z <- 3 == 2
z
```

```
## [1] FALSE
```

```
class(z)
```

```
## [1] "logical"
```

Here the == is a relational operator asking if 3 is equal to 2. Remember that in R, if you just use one = when you actually assign a value. You can see the other *relational operators* by typing

```
?Comparison
```

In future sections you will see how useful relational operators can be.

Advanced: Mathematically, the values in `pop` are integers and there is an integer class in R. However, by default, numbers are assigned class numeric even when they are round integers, for example `class(1)` returns numeric. You can turn them into class integer with `as.integer()` or by adding an L like this: `1L`. Note the class by typing: `class(1L)`.

Factors

In the `murders` dataset we might expect the region to also be a character vector. However, it is not:

```
class(murders$region)
```

```
## [1] "factor"
```

it is a *factor*. Factors are useful for storing categorical data. Notice that there are only 4 regions:

```
levels(murders$region)
```

```
## [1] "Northeast"      "South"           "North Central"  "West"
```

So, in the background, R stores these *levels* as integers and keeps a map to keep track of the labels. This is more memory efficient than storing all the characters. However, factors are also a source of confusion as they can easily be confused with characters but behave differently in different contexts. We will see more of this later.

In general, we recommend avoiding factors as much as possible although they are sometimes necessary to fit models containing categorical data.

Lists

Data frames are a special case of *lists*. We will cover lists in more detail later but know that they are useful because you can store any combination of other types. Here is an example of a list we created for you:

```
record <- list(name = "John Doe",
               student_id = 1234,
               grades = c(95, 82, 91, 97, 93),
               final_grade = "A")
```

```
record
```

```
## $name
## [1] "John Doe"
##
## $student_id
## [1] 1234
##
## $grades
## [1] 95 82 91 97 93
##
## $final_grade
## [1] "A"
```

```
class(record)
```

```
## [1] "list"
```

We won't be using lists until later but you might encounter one in your own exploration of R. Note that, as with data frames, you can extract the components with the accessor `$`. In fact, data frames are a type of list.

```
record$student_id
```

```
## [1] 1234
```

We can also use double brackets like this:

```
record[["student_id"]]
```

```
## [1] 1234
```

You should get used to the fact that in R there are several ways to do the same thing, in particular accessing entries.