HTTP

INTERMEDIATE IMPORTING DATA IN R

Filip Schouwenaars
Instructor, DataCamp

datacamp

Data on the web

- Already worked with it!

- Many packages handle it for you

- File formats useful for web technology

- JSON

HTTP

- HyperText Transfer Protocol

- Rules about data exchange between computers
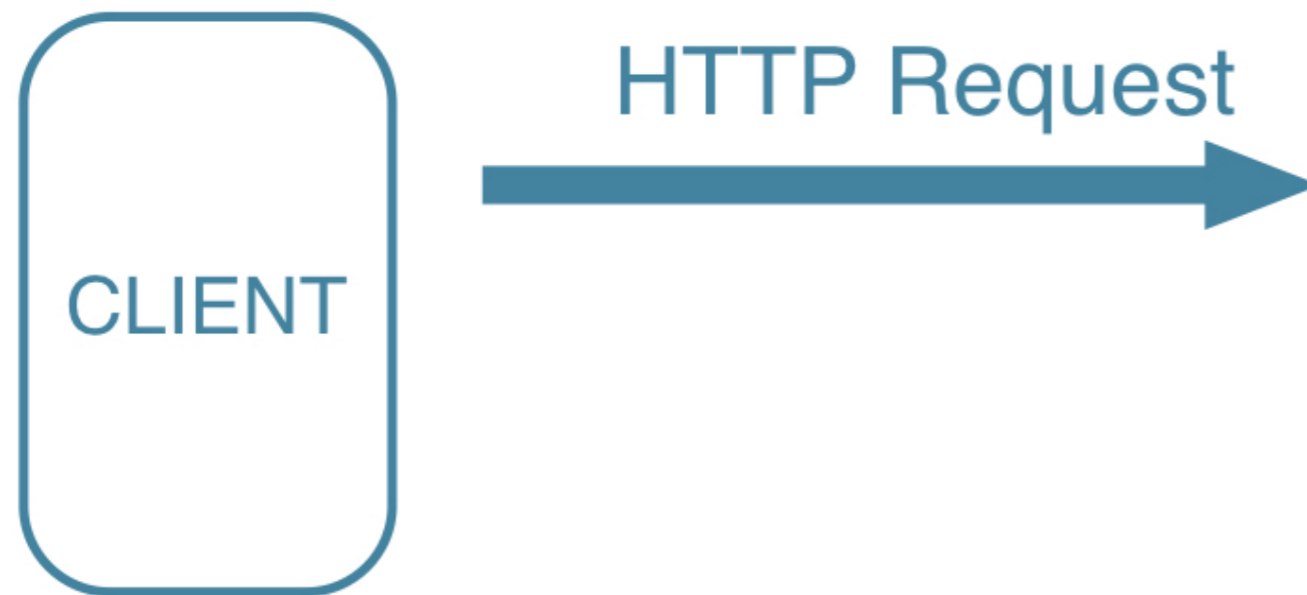
- Language of the web

CLIENT

HTTP

- HyperText Transfer Protocol

- Rules about data exchange between computers

- Language of the web

HTTP

- HyperText Transfer Protocol

- Rules about data exchange between computers
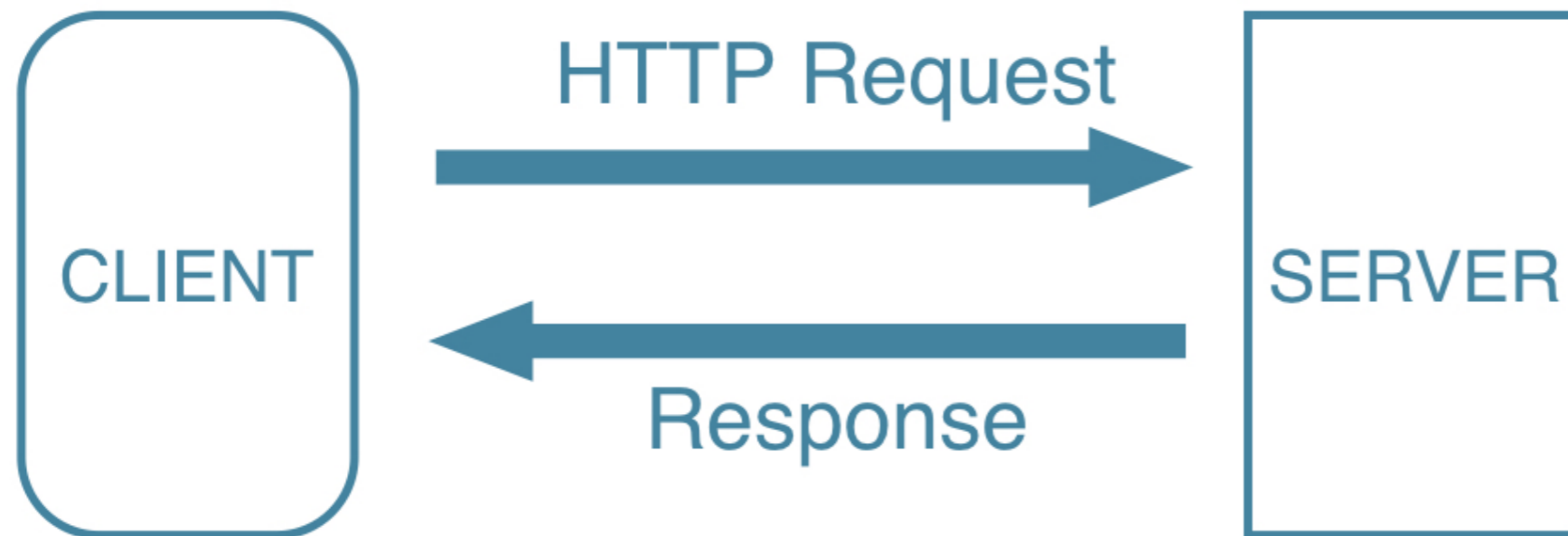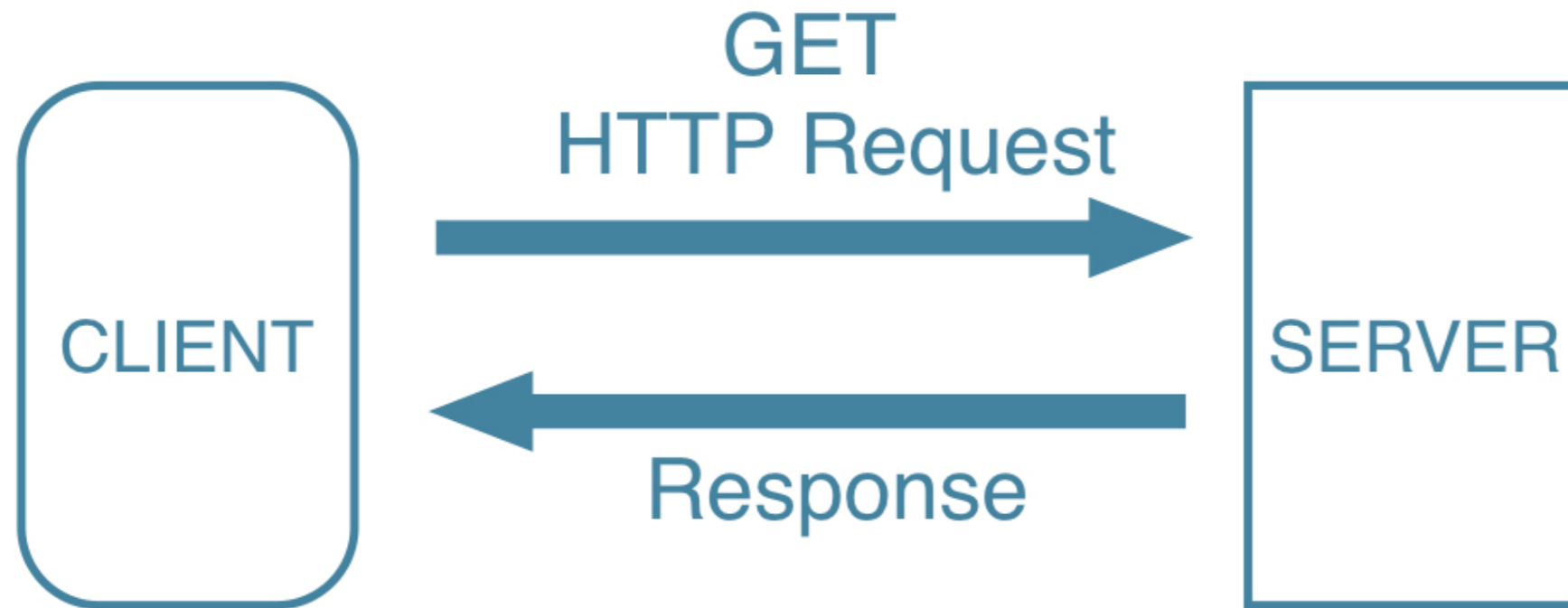
- Language of the web

HTTP

- HyperText Transfer Protocol

- Rules about data exchange between computers

- Language of the web

HTTP

- HyperText Transfer Protocol

- Rules about data exchange between computers

- Language of the web

Example: CSV

http://s3.amazonaws.com/ ... /states.csv

```r
# Manually download file through web browser


read.csv(url("path/to/states.csv"))
```

```
        state      capital pop_mill area_sqm
1 South Dakota      Pierre    0.853    77116
2    New York       Albany   19.746    54555
3      Oregon        Salem    3.970    98381
4     Vermont   Montpelier    0.627     9616
5      Hawaii     Honolulu    1.420    10931
```

Example: CSV

```
read.csv("http://s3.amazonaws.com/
              assets.datacamp.com/course/
              importing_data_into_r/states.csv")
```

```
          state      capital pop_mill area_sqm
1 South Dakota       Pierre    0.853    77116
2     New York       Albany   19.746    54555
3       Oregon        Salem    3.970    98381
4      Vermont   Montpelier    0.627     9616
5       Hawaii     Honolulu    1.420    10931
```

- R sees it's a URL, does GET request, and reads in the CSV file

Example: CSV

```r
read.csv("https://s3.amazonaws.com/
                assets.datacamp.com/course/
                importing_data_into_r/states.csv")
```

```
        state    capital pop_mill area_sqm
1 South Dakota     Pierre    0.853    77116
2     New York     Albany   19.746    54555
3       Oregon      Salem    3.970    98381
4      Vermont Montpelier    0.627     9616
5       Hawaii   Honolulu    1.420    10931
```

- HTTPS support since R version 3.2.2

Let's practice!

INTERMEDIATE IMPORTING DATA IN R

Downloading files

INTERMEDIATE IMPORTING DATA IN R

Filip Schouwenaars
Instructor, DataCamp

datacamp

Example: Excel

```r
library(readxl)
read_excel("http://s3.amazonaws.com/
               assets.datacamp.com/course/
               importing_data_into_r/cities.xlsx")
```

```
Error:
'http://s3.amazonaws.com/assets.datacamp.com/course/importing_data_into_r/cities.xlsx'
does not exist in current working directory.
```

download.file()

```r
url <- "http://s3.amazonaws.com/assets.datacamp.com/
         course/importing_data_into_r/cities.xlsx"
dest_path <- file.path("~", "local_cities.xlsx")
download.file(url, dest_path)
```

```
// Messages showing download progress omitted
```

```r
read_excel(dest_path)
```

```
     Capital Population
1  New York   16044000
2     Berlin    3433695
3     Madrid    3010492
4  Stockholm    1683713
```

Why download.file()?

- Reproducibility

- HTTP from inside R
  - Authentication

  - Additional parameters

  - httr - Hadley Wickham

Let's practice!

INTERMEDIATE IMPORTING DATA IN R