

Introduction to R Programming

Slide Set 10: Wider R Functionality

Maria Ptashkina

Barcelona GSE ITFD

September 2021

Table of Contents

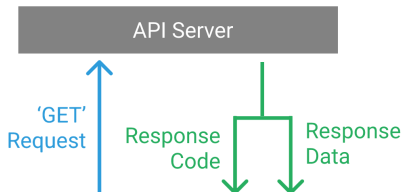
- 1 APIs
- 2 Web Scraping
- 3 Text as Data
- 4 Geospatial Data
- 5 R Markdown

What is an API

- APIs (Application Programming Interface) is becoming a more common way to get interesting data
- Many large websites like Reddit, Twitter and Facebook offer APIs, but recently international organizations and government agencies (World Bank, UN Comtrade, Bureau of Economic Analysis, etc.) are increasingly offering APIs
- 'API' is a general term for the place where one computer program interacts with another, or with itself

Making Requests

- We will be working with web APIs, where two different computers - a client and server - interact with each other to request and provide data
- When a website set up an API, they are essentially setting up a computer that waits for data requests
- Once the computer receives a data request, it will do its own processing of the data and send it to the computer that requested it



Example 1: GDP Data from BEA

- In this exercise we will take quarterly GDP data from the Bureau of Economic Analysis, this example shows a relatively complicated way to work with APIs
- You first need to do a simple and quick registration at BEA website
▶ Register to get the API key
- Every API provides a detailed documentation about how to use it
▶ Documentation
- For GDP in particular BEA publishes a primer ▶ Primer
- To do a request we need to know the API address and parameters to get (this information can be found in the API documentation)
- Often you would find the data returned to use massive - use R data wrangling packages to get exactly what you want

Example 2. Trade Data from Comtrade

- In this exercise we will get the data from UN Comtrade using a user defined function that they provide [▶ Comtrade API](#)
- Note that some APIs have very strict usage limits: for example, UN Comtrade limits you to 5 codes for importer, exporter and time period per request. You could overcome that with writing a loop, but if you hit 100 requests per hour, the server will automatically ban you
- Bottom line: pay attention to API usage limits!

Example 3. World Bank Data

- Some data providers have a specific R package written to easily access their APIs, and make the user experience exquisite
- This is the case, for example, for the World Bank data, which is called `wbstats` and is very nice and easy to use

- `quantmod` is a powerful package, which contains a helper function called `getSymbols()`
- It goes to a site and pulls back a data series for a particular stock or economic indicator
- The package supports a variety of sources, including Google Finance, Yahoo Finance, and FRED

Table of Contents

- ① APIs
- ② Web Scraping
- ③ Text as Data
- ④ Geospatial Data
- ⑤ R Markdown

What is Web Scraping

- Web scraping is a technique for converting the data present in unstructured format (HTML tags) over the web to the structured format which can easily be accessed and used
- While API gives you a direct access to the data you want, web scraping is usually much messier
- You can use web scraping when there is no API access to the data you want or API is too limited or expensive

A Quick Look at HTML and XML

- Any webpage you visit has a particular, expected general structure, which usually consists of two types of code: HTML and XML
- HTML / XML is not a programming language, it's a *markup language* — it describes the content and structure of a web page
- HTML code has an expected format and structure, to make it easier for people to develop web pages
- XML looks very similar to HTML, but users can create their own named tags

HTML

```
<html>
<head>
<title>Page Title</title>
</head>
<body>
<h1>This is a Heading</h1>
<p>This is a paragraph.</p>
</body>
</html>
```

XML

```
<note>
  <to>Maria</to>
  <from>BGSE</from>
  <heading>Intro to R</heading>
  <body>Great job!</body>
</note>
```

Web Scraping Using R

- To mine web data, it's important that you can see the underlying code and understand how it relates to what you are seeing on the page
- The best way to do this is to use the Developer Tools that come with Google Chrome (press Ctrl+Shift+C in Windows or Cmd+Option+C on a Mac)
- We can mine elements of a web page and treat them like lists in R
- The `rvest` library is a part of `tidyverse`, and lets users easily scrape ('harvest') data from web pages

Table of Contents

- 1 APIs
- 2 Web Scraping
- 3 Text as Data
- 4 Geospatial Data
- 5 R Markdown

- For social scientists, the information encoded in text is a rich complement to the more structured kinds of data traditionally used in research
- In recent years have seen an explosion of empirical economics research using text as data
 - In finance, text from financial news, social media, and company filings is used to predict asset price movements and study the causal impact of new information
 - In macroeconomics, text is used to forecast variation in inflation and unemployment, and estimate the effects of policy uncertainty
 - In media economics, text from news and social media is used to study the drivers and effects of political slant
 - In political economy, text from politicians' speeches is used to study the dynamics of political agendas and debate
- Text mining and natural language processing applications range from analyzing word frequencies, to sentiment analysis, to topic modeling...
- This is a large topic, so we will look only at a tiny bit of code + check out the tutorials

Table of Contents

- ① APIs
- ② Web Scraping
- ③ Text as Data
- ④ Geospatial Data
- ⑤ R Markdown

Geospatial Data

- The use of geospatial data – data that can be mapped using geographic information systems (GIS) – has become increasingly widespread in the social sciences
- Applications not only extend to the analysis of classical geographical entities (e.g. countries) but increasingly also to analyses of micro-level data, including respondent information from georeferenced surveys
- Geospatial data comprise information about geometries (points, lines, polygons, grids) related to a location on a map
- This relation is usually established through geo-coordinates, which carry information on longitude and latitude
- There are many packages and instruments in R which allow you to work with geospatial data
- I will show you only simple map visualization, but there is so much more you can do - check out the tutorials!

Table of Contents

- 1 APIs
- 2 Web Scraping
- 3 Text as Data
- 4 Geospatial Data
- 5 R Markdown

R Markdown and Notebooks

- R Markdown is an open-source tool for producing reproducible reports in R
- It enables you to keep all of your code, results, plots, and writing in one place
- R Markdown allows to generate a report (most of the time in PDF, HTML, Word or as a beamer presentation) that is automatically generated from a file written within RStudio
- There are also other tools, such as Jupiter Notebook (used for Python, but which can support codes in other languages, and even Stata codes)

References and Resources

- Data Acquisition in R [▶ Tutorial](#)

API

- R API Tutorial: Getting Started with APIs in R [▶ Tutorial](#)
- GDP Data via API [▶ Tutorial](#)
- Using the UN Comtrade data API with R [▶ Data](#)
- The World Bank data API [▶ Tutorial](#)
- Data Acquisition in R [▶ Tutorial](#)

Web Scraping

- A simple example of webscraping with R : the top 5% economist ranking [▶ Tutorial](#)
- Tutorial: Web Scraping in R with rvest [▶ Tutorial](#)
- Tidy web scraping in R — Tutorial and resources [▶ Tutorial](#)
- Yahoo Finance Web Scraping with R [▶ Tutorial](#)

References and Resources

Text as Data

- Gentzkow (2019) [▶ Paper](#)
- Text Mining with R [▶ Book](#)
- An Introduction to Text Processing and Analysis with R [▶ Tutorial](#)
- Basic Text Processing in R
- An Introduction to Text Processing and Analysis with R [▶ Tutorial](#)
- Text Mining Twitter Data With TidyText in R [▶ Tutorial](#)

Geospatial Data

- Using Geospatial Data in R [▶ Tutorial](#)
- Chapter 6 Maps [▶ Tutorial](#)
- Geocomputation with R [▶ Tutorial](#)

R Markdown

- Learn R Markdown [▶ Tutorial](#)
- R Markdown [▶ Tutorial](#)
- Getting Started with R Markdown - Guide and Cheatsheet [▶ Tutorial](#)
- Getting started in R markdown [▶ Tutorial](#)
- R for Data Science [▶ Tutorial](#)