



Web Scraping

Aplicaciones Financieras



Richard P. Pérez Palma Ponce
rperezpalma19@gmail.com

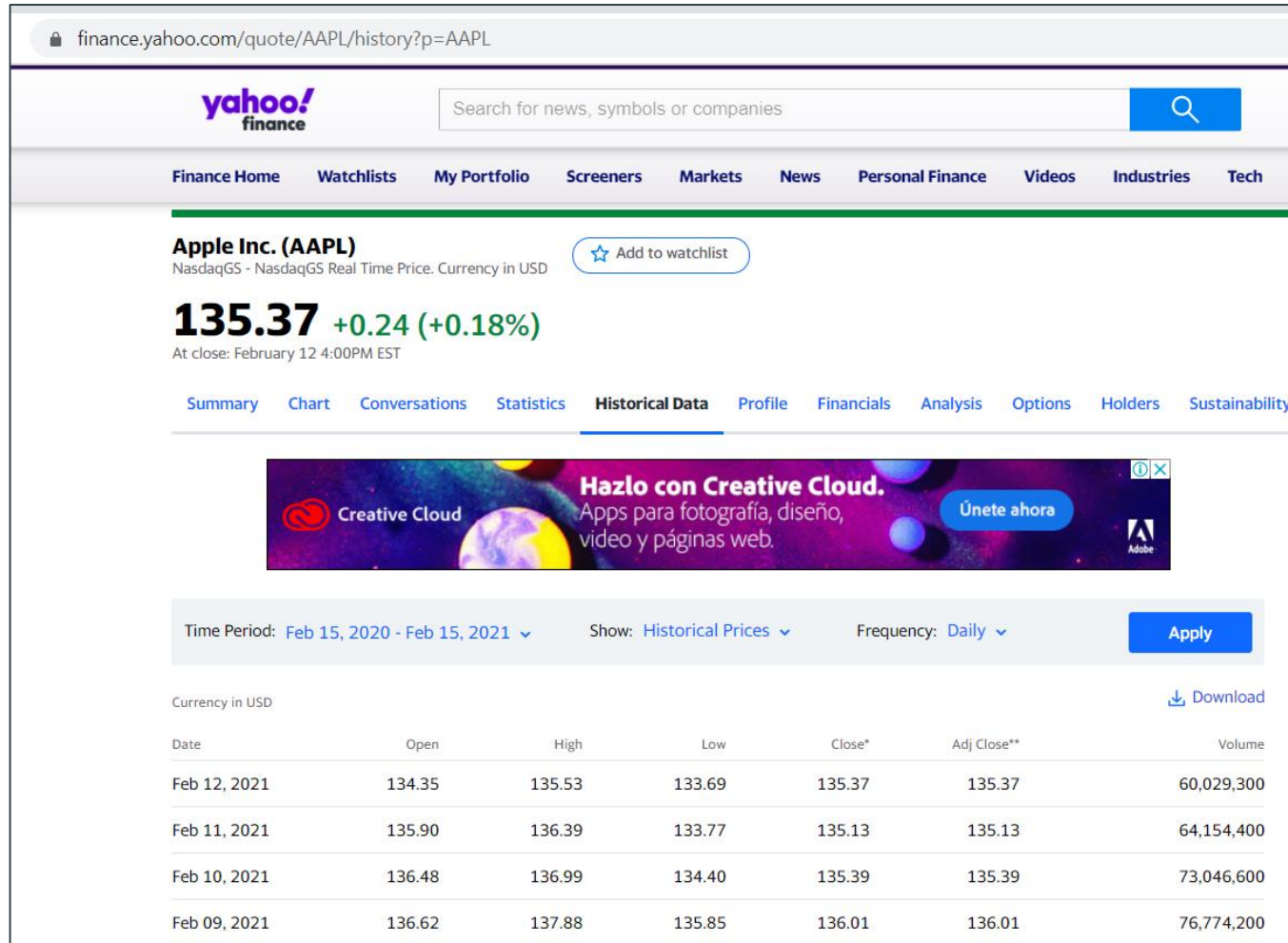
Método 1

Librerías en R
(quantmod)

Web Scraping: Librería en R

Yahoo Finance

<https://finance.yahoo.com/>



finance.yahoo.com/quote/AAPL/history?p=AAPL

yahoo!
finance

Search for news, symbols or companies

Finance Home Watchlists My Portfolio Screeners Markets News Personal Finance Videos Industries Tech

Apple Inc. (AAPL)
NasdaqGS - NasdaqGS Real Time Price. Currency in USD [Add to watchlist](#)

135.37 +0.24 (+0.18%)
At close: February 12 4:00PM EST

Summary Chart Conversations Statistics **Historical Data** Profile Financials Analysis Options Holders Sustainability

Hazlo con Creative Cloud.
Apps para fotografía, diseño, video y páginas web. [Únete ahora](#)

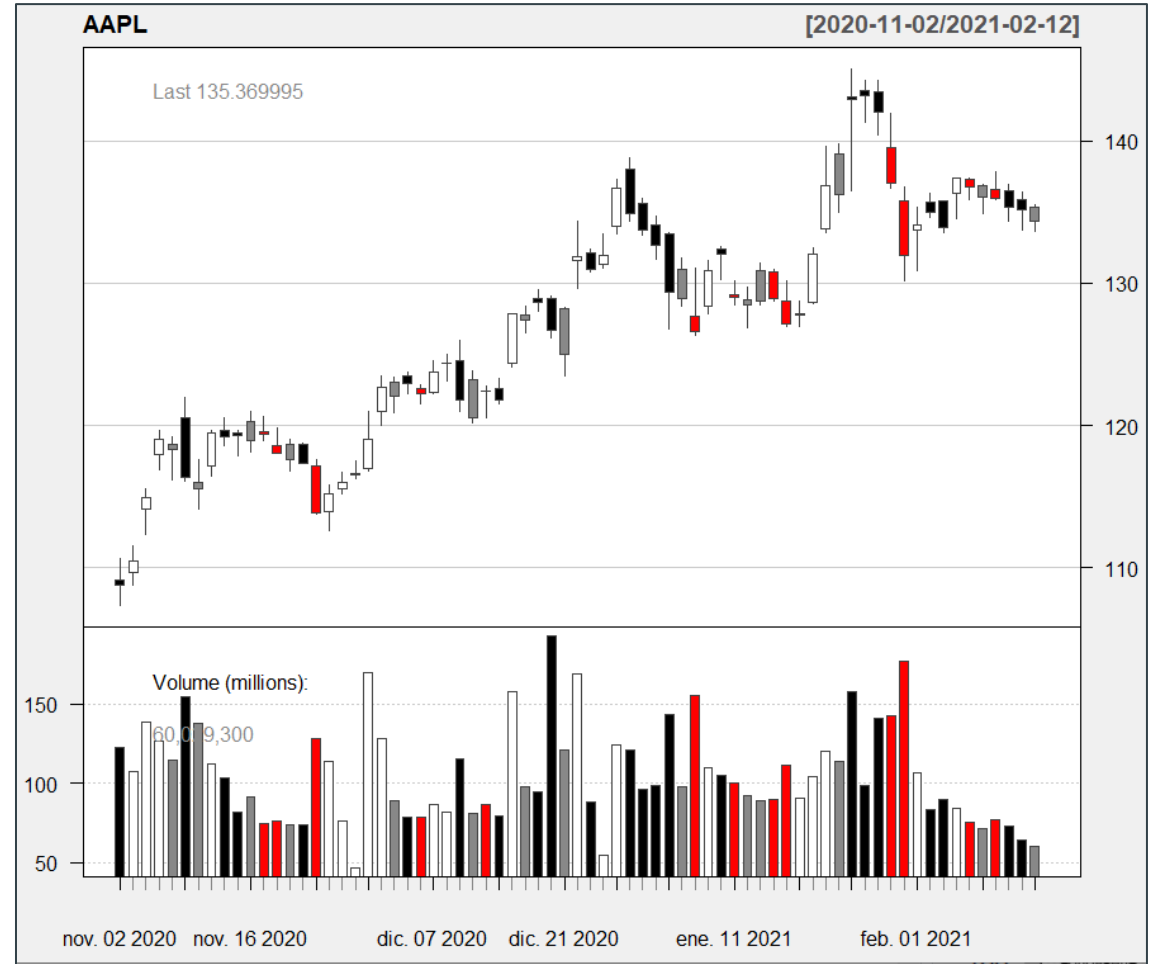
Time Period: Feb 15, 2020 - Feb 15, 2021 Show: Historical Prices Frequency: Daily [Apply](#)

Currency in USD [Download](#)

Date	Open	High	Low	Close*	Adj Close**	Volume
Feb 12, 2021	134.35	135.53	133.69	135.37	135.37	60,029,300
Feb 11, 2021	135.90	136.39	133.77	135.13	135.13	64,154,400
Feb 10, 2021	136.48	136.99	134.40	135.39	135.39	73,046,600
Feb 09, 2021	136.62	137.88	135.85	136.01	136.01	76,774,200

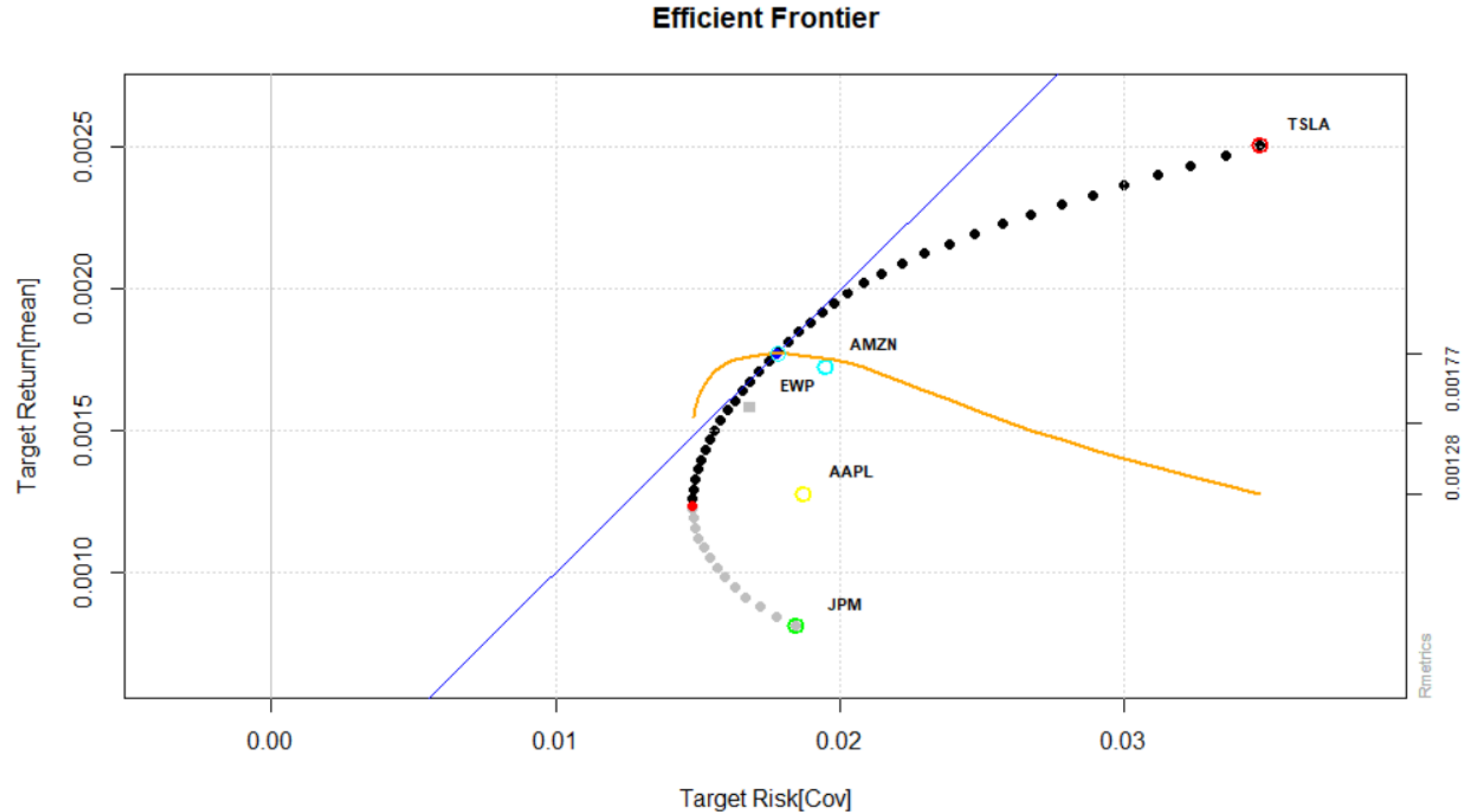
Web Scraping: Librería en R

Yahoo Finance (*quantmod*)



Web Scraping: Librería en R

Yahoo Finance (*fPortfolio*)




Web Scraping: Librería en R

Banco Mundial (*wbstats*)

data.worldbank.org

Learn how the World Bank Group is helping countries with COVID-19 (coronavirus). [Find Out >](#)

 **THE WORLD BANK** | [Data](#)

This page in: [English](#) [Español](#) [Français](#) [العربية](#) [中文](#)

New to this site? [Start Here](#)

[DataBank](#) [Microdata](#) [Data Catalog](#)

World Bank Open Data

Free and open access to global development data

Search data e.g. GDP, population, Indonesia

Browse by [Country](#) or [Indicator](#)

MOST RECENT

Catastrophic expenditure for health: from antiquity to today [↗](#)

Mark G Shrimme, Feb 19, 2021

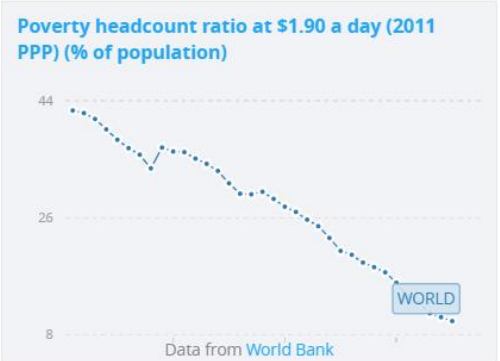
Supporting water and disaster risk management in Mozambique using Google Earth Engine [↗](#)

Sun Hwa Song, Siddhesh Kaushik, Leslie Ricketts, Feb 18, 2021

Ensuring continuity in survey operations

WHAT YOU CAN LEARN WITH OPEN DATA [<](#) [>](#)

Poverty headcount ratio at \$1.90 a day (2011 PPP) (% of population)



44
26
8


WORLD

Data from [World Bank](#)

Extreme Poverty

Atlas of Sustainable Development Goals 2020

From World Development Indicators



Web Scraping: Librería en R

Bloomberg (*Rblpapi*)



Método 2

APIs

(XML - JSONLITE)

Web Scraping: APIs

BANCO CENTRAL DE RESERVA DEL PERÚ (BCRP)

<https://estadisticas.bcrp.gob.pe/estadisticas/series/>



The screenshot shows the BCRPData website interface. The left sidebar contains a menu with the following items: 'Consulta de Series' (with sub-items 'Por Categoría', 'Por Frecuencia', 'Por Series'), 'Series Pre-Definidas' (with sub-items 'Nota Semanal Mensual', 'Nota Semanal Trimestral', 'Cuadros Anuales'), 'Herramientas' (with sub-items 'Listas Personalizadas', 'Add-In para Excel', 'API para Desarrolladores' (highlighted with a red box), 'App para móviles', 'Metadatos'), and 'Encuesta de Expectativas'. The main content area displays 'CATEGORÍAS (13807)' with sub-sections: 'Moneda y Crédito (4265)', 'Precios (127)', 'Tipo de Cambio (135)', 'Balanza de Pagos (1062)', and 'PBI y Mercado Laboral (707)'. A large red arrow points from the 'API para Desarrolladores' link in the sidebar to the 'Tipo de Cambio' section in the main content area.

Gráfico: Esta opción mostrará un gráfico en formato flash.

<https://estadisticas.bcrp.gob.pe/estadisticas/series/api/PN01288PM/grafico>

<https://estadisticas.bcrp.gob.pe/estadisticas/series/api/PN01288PM-PN01289PM/grafico>

XLS: Esta opción generará un archivo en formato XLS.

<https://estadisticas.bcrp.gob.pe/estadisticas/series/api/PN01288PM/xls>

<https://estadisticas.bcrp.gob.pe/estadisticas/series/api/PN01288PM-PN01289PM/xls/2013-1/2016-9>

XML: Esta opción mostrará un archivo en formato XML.

<https://estadisticas.bcrp.gob.pe/estadisticas/series/api/PN01288PM/xml>

<https://estadisticas.bcrp.gob.pe/estadisticas/series/api/PN01288PM-PN01289PM/xml/2013-1/2016-9>

JSON y JSONP: Estas opciones mostrarán salidas en formato JSON (sin paréntesis) y JSONP (con paréntesis).

<https://estadisticas.bcrp.gob.pe/estadisticas/series/api/PN01288PM/json>

<https://estadisticas.bcrp.gob.pe/estadisticas/series/api/PN01288PM-PN01289PM/json/2013-1/2016-9>

<https://estadisticas.bcrp.gob.pe/estadisticas/series/api/PN01288PM/jsonp>

<https://estadisticas.bcrp.gob.pe/estadisticas/series/api/PN01288PM-PN01289PM/jsonp/2013-1/2016-9>

Web Scraping: APIs

BCRP (XML y JSON)

XML

```
← → ↺ estadisticas.bcrp.gob.pe/estadisticas/series/api/PN01770AM/xml/2000-1/2020-12

This XML file does not appear to have any style information associated with it. The document tree is shown below.

<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<chart>
  <config title="PBI">
    <serie name="Producto bruto interno y demanda interna (índice 2007=100) - PBI" dec="1"/>
  </config>
  <periods>
    <period name="Ene.03">
      <v>73.2223699646929</v>
    </period>
    <period name="Feb.03">
      <v>71.0794325012595</v>
    </period>
    <period name="Mar.03">
      <v>74.3427233378699</v>
    </period>
    <period name="Abr.03">
      <v>80.6677117569535</v>
    </period>
    <period name="May.03">
      <v>83.2226461043316</v>
    </period>
    <period name="Jun.03">
      <v>80.8537747227157</v>
    </period>
    <period name="Jul.03">
      <v>79.2070173867754</v>
    </period>
    <period name="Ago.03">
      <v>74.2120140864971</v>
    </period>
    <period name="Sep.03">
      <v>73.8678355991462</v>
    </period>
    <period name="Oct.03">
      <v>73.8678355991462</v>
    </period>
    <period name="Nov.03">
      <v>73.8678355991462</v>
    </period>
    <period name="Dic.03">
      <v>73.8678355991462</v>
    </period>
    <period name="Ene.04">
      <v>75.9132671626346</v>
    </period>
    <period name="Feb.04">
      <v>74.0974361472172</v>
    </period>
    <period name="Mar.04">
      <v>86.5105704991306</v>
    </period>
    <period name="Jun.04">
      <v>83.2874592708797</v>
    </period>
    <period name="Jul.04">
      <v>77.4025940698837</v>
    </period>
    <period name="Oct.04">
      <v>79.6324554115629</v>
    </period>
    <period name="Nov.04">
      <v>79.9791613436588</v>
    </period>
    <period name="Feb.05">
      <v>80.134460157334</v>
    </period>
    <period name="Mar.05">
      <v>92.1414553918693</v>
    </period>
    <period name="Jun.05">
      <v>88.4557732759908</v>
    </period>
    <period name="Jul.05">
      <v>82.0878667862381</v>
    </period>
    <period name="Oct.05">
      <v>84.8027241979274</v>
    </period>
    <period name="Nov.05">
      <v>85.6590427598994</v>
    </period>
    <period name="Feb.06">
      <v>84.6285223502132</v>
    </period>
    <period name="Mar.06">
      <v>97.7619066472792</v>
    </period>
    <period name="Jun.06">
      <v>95.0729001070201</v>
    </period>
    <period name="Jul.06">
      <v>88.7693298777215</v>
    </period>
    <period name="Oct.06">
      <v>92.1822069581638</v>
    </period>
    <period name="Nov.06">
      <v>89.948891875024</v>
    </period>
    <period name="Feb.07">
      <v>88.6754172656189</v>
    </period>
    <period name="Mar.07">
      <v>104.744753291331</v>
    </period>
    <period name="Jun.07">
      <v>101.255860662521</v>
    </period>
    <period name="Jul.07">
      <v>100.225943716493</v>
    </period>
    <period name="Oct.07">
      <v>103.154829161475</v>
    </period>
    <period name="Nov.07">
      <v>98.5149773291643</v>
    </period>
    <period name="Feb.08">
      <v>100.709860257349</v>
    </period>
    <period name="Mar.08">
      <v>112.120913726693</v>
    </period>
    <period name="Jun.08">
      <v>112.211609963734</v>
    </period>
    <period name="Jul.08">
      <v>110.725068248671</v>
    </period>
    <period name="Oct.08">
      <v>111.641043578987</v>
    </period>
    <period name="Nov.08">
      <v>103.013101977079</v>
    </period>
    <period name="Feb.09">
      <v>101.019865758232</v>
    </period>
    <period name="Mar.09">
      <v>114.22464786173</v>
    </period>
    <period name="Jun.09">
      <v>108.887627134331</v>
    </period>
    <period name="Jul.09">
      <v>110.86582898004</v>
    </period>
    <period name="Oct.09">
      <v>112.95597594939</v>
    </period>
    <period name="Nov.09">
      <v>106.153025639292</v>
    </period>
    <period name="Feb.10">
      <v>106.146322827933</v>
    </period>
    <period name="Mar.10">
      <v>123.028902592226</v>
    </period>
    <period name="Jun.10">
      <v>123.162705739863</v>
    </period>
    <period name="Jul.10">
      <v>122.292317240048</v>
    </period>
    <period name="Oct.10">
      <v>123.835093598882</v>
    </period>
    <period name="Nov.10">
      <v>122.292317240048</v>
    </period>
  </periods>
</chart>
```

JSON

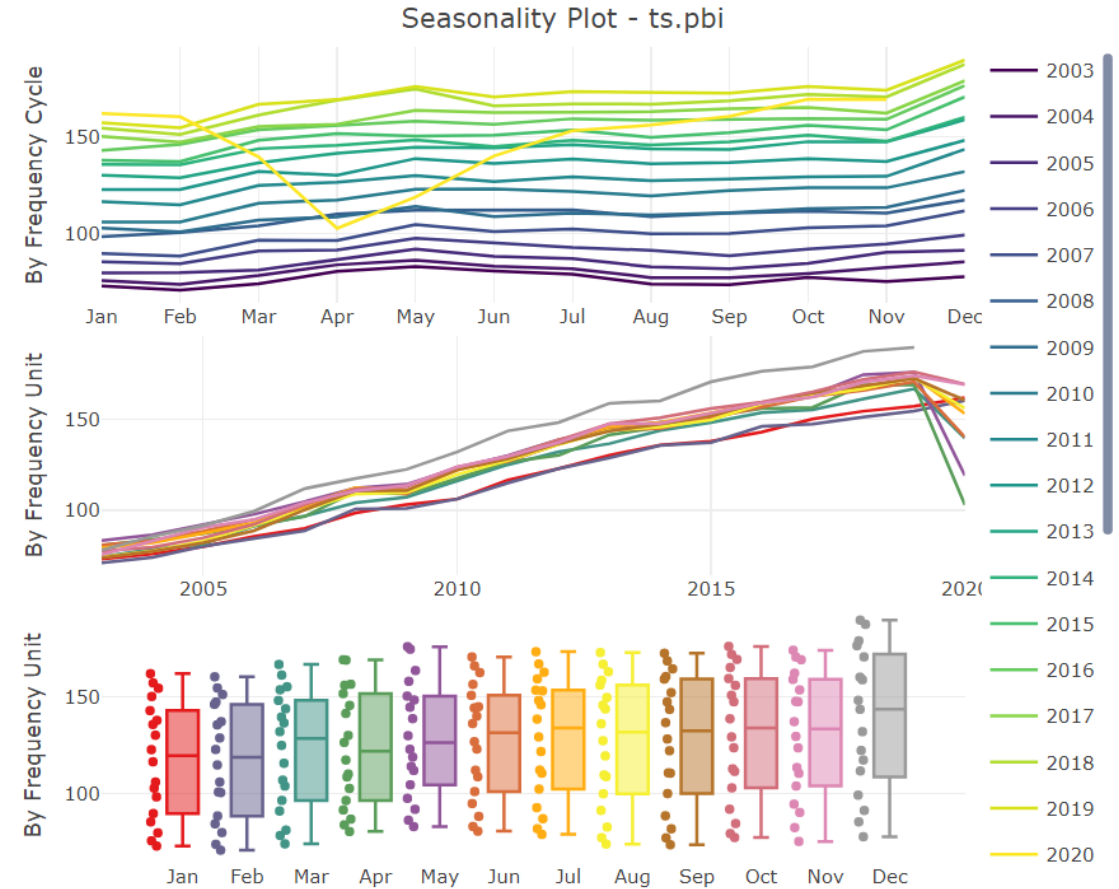
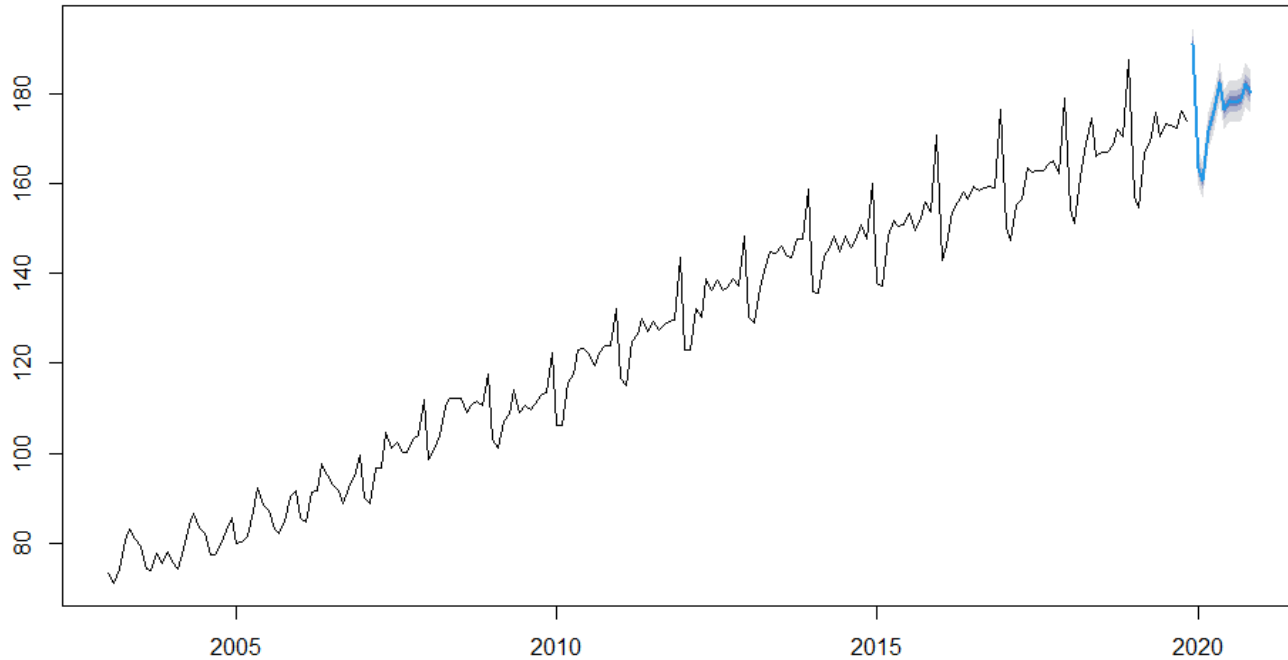
```
← → ↺ estadisticas.bcrp.gob.pe/estadisticas/series/api/PN01770AM/json/2000-1/2020-12

{ "config": { "title": "Producto bruto interno y demanda interna (índice 2007=100)", "series": [ { "name": "Ene.2003", "values": [73.2223699646929] }, { "name": "Feb.2003", "values": [71.0794325012595] }, { "name": "Mar.2003", "values": [74.3427233378699] }, { "name": "Jun.2003", "values": [80.8537747227157] }, { "name": "Jul.2003", "values": [79.2070173867754] }, { "name": "Oct.2003", "values": [73.8678355991462] }, { "name": "Nov.2003", "values": [73.8678355991462] }, { "name": "Dic.2003", "values": [73.8678355991462] }, { "name": "Ene.2004", "values": [75.9132671626346] }, { "name": "Feb.2004", "values": [74.0974361472172] }, { "name": "Mar.2004", "values": [86.5105704991306] }, { "name": "Jun.2004", "values": [83.2874592708797] }, { "name": "Jul.2004", "values": [77.4025940698837] }, { "name": "Oct.2004", "values": [79.6324554115629] }, { "name": "Nov.2004", "values": [79.9791613436588] }, { "name": "Feb.2005", "values": [80.134460157334] }, { "name": "Mar.2005", "values": [92.1414553918693] }, { "name": "Jun.2005", "values": [88.4557732759908] }, { "name": "Jul.2005", "values": [82.0878667862381] }, { "name": "Oct.2005", "values": [84.8027241979274] }, { "name": "Nov.2005", "values": [85.6590427598994] }, { "name": "Feb.2006", "values": [84.6285223502132] }, { "name": "Mar.2006", "values": [97.7619066472792] }, { "name": "Jun.2006", "values": [95.0729001070201] }, { "name": "Jul.2006", "values": [88.7693298777215] }, { "name": "Oct.2006", "values": [92.1822069581638] }, { "name": "Nov.2006", "values": [89.948891875024] }, { "name": "Feb.2007", "values": [88.6754172656189] }, { "name": "Mar.2007", "values": [104.744753291331] }, { "name": "Jun.2007", "values": [101.255860662521] }, { "name": "Jul.2007", "values": [100.225943716493] }, { "name": "Oct.2007", "values": [103.154829161475] }, { "name": "Nov.2007", "values": [98.5149773291643] }, { "name": "Feb.2008", "values": [100.709860257349] }, { "name": "Mar.2008", "values": [112.120913726693] }, { "name": "Jun.2008", "values": [112.211609963734] }, { "name": "Jul.2008", "values": [110.725068248671] }, { "name": "Oct.2008", "values": [111.641043578987] }, { "name": "Nov.2008", "values": [103.013101977079] }, { "name": "Feb.2009", "values": [101.019865758232] }, { "name": "Mar.2009", "values": [114.22464786173] }, { "name": "Jun.2009", "values": [108.887627134331] }, { "name": "Jul.2009", "values": [110.86582898004] }, { "name": "Oct.2009", "values": [112.95597594939] }, { "name": "Nov.2009", "values": [106.153025639292] }, { "name": "Feb.2010", "values": [106.146322827933] }, { "name": "Mar.2010", "values": [123.028902592226] }, { "name": "Jun.2010", "values": [123.162705739863] }, { "name": "Jul.2010", "values": [122.292317240048] }, { "name": "Oct.2010", "values": [123.835093598882] }, { "name": "Nov.2010", "values": [122.292317240048] } ] } }
```

Web Scraping: APIs

BCRP (forecast)

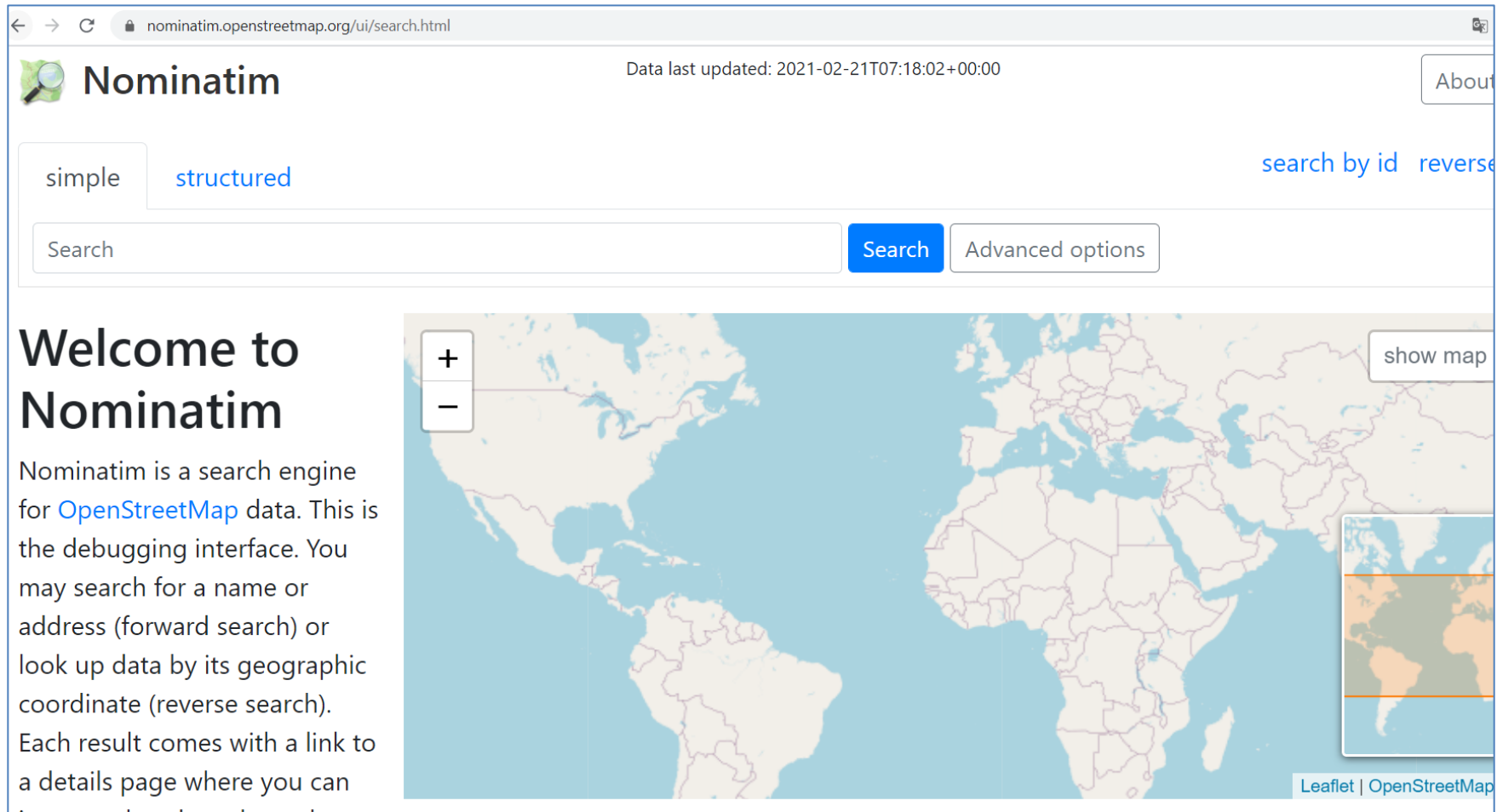
Forecasts from ARIMA(1,0,1)(0,1,2)[12] with drift



Web Scraping: APIs

Nominatim - Openstreetmap

<https://nominatim.openstreetmap.org/search?format=xml&q=Congreso%20de%20la%20Republica,%20PERU>



Método 3

Web Scrapping Estático
(*rvest*)

Web Scraping: Página Web Estática

Investing

<https://es.investing.com/indices/major-indices>

The screenshot displays the Investing.com website's major indices page. The browser's address bar shows the URL `es.investing.com/indices/major-indices`. The page header includes a search bar and a login/register prompt. Below the header, a table titled "Principales índices bursátiles" lists various stock indices with columns for the index name, last price, high, low, change, percentage change, and time. A tooltip for the table shows its dimensions as 652 x 1047.2. To the right of the table, there is a "Descargar datos" button. Below the table, a small chart for the IBEX 35 index is visible. The Chrome DevTools overlay on the right shows the "Elements" panel with the HTML structure of the page. The table's HTML is highlighted, showing the `<table>` tag with attributes `class="genTbl closedTbl elpTbl elp20 crossRatesTbl"`. The table's structure includes a `<thead>` with columns for index name, high, low, change, percentage change, and time, and a `<tbody>` with data rows. The console panel at the bottom shows a message about the Chrome 88 update.

Principales índices bursátiles

Índice	Último	Máximo	Mínimo	Var.	% Var.	Hora
IBEX 35	8.164,00	8.167,00	8.105,00	+109,00	+1,35%	10:19:38
I.G. Bolsa Madrid	816,75	817,66	813,39	+9,47	+1,17%	10:04:00
Dow Jones	31.458,40	31.475,31	31.347,80	+27,70	+0,09%	12/02
S&P 500	3.934,83	3.937,23	3.905,78	+18,45	+0,47%	12/02
Nasdaq	14.095,47	14.102,04	13.937,71	+69,70	+0,50%	12/02
Russell 2000	2.287,90	2.293,73	2.264,38	+2,58	+0,11%	12/02
S&P 500 VIX	19,97	22,45	19,95	0,00	0,00%	12/02
S&P/TSX	18.460,21	18.460,79	18.335,83	+67,22	+0,37%	12/02
Bovespa	119.428,72	119.763,36	118.162,68	+128,89	+0,11%	12/02
S&P Merval	51.466,21	52.104,65	51.245,58	-188,35	-0,37%	12/02
COLCAP	1.365,95	1.368,88	1.360,43	-1,84	-0,13%	12/02
S&P Lima Gene...	22.413,98	22.415,82	22.147,29	+262,04	+1,18%	12/02
S&P CLX IPSA	4.599,09	4.599,71	4.531,61	+50,89	+1,12%	12/02
S&P/BMV IPC	44.202,52	44.239,79	43.923,87	+142,43	+0,32%	12/02
DAX	14.104,15	14.140,15	14.083,60	+54,26	+0,38%	10:10:42
FTSE 100	6.683,95	6.685,65	6.632,56	+94,16	+1,42%	10:10:42
CAC 40	5.768,95	5.769,72	5.722,16	+65,28	+1,14%	10:10:42

Descargar datos

Índices Acciones

IBEX 35 8.164,00

Futuros S&P 500 3.947,60

Desplome del Dogecoin y otras cripto tras el 'aviso' de venta de Elon Musk

Web Scraping: Página Web Estática

Gestión

<https://gestion.pe/noticias/afp/1/>

The screenshot shows a web browser at the URL <https://gestion.pe/noticias/afp/1/>. The page features a dark header with the 'GESTIÓN' logo, a 'Suscríbete' button, and a user profile icon. Below the header, there's a navigation bar with links: 'HOY INTERESA', 'Últimas Noticias', 'Coronavirus Perú', 'Afiliados A AFP', and 'Tipo De Camb'. The main content area has a large promotional banner for 'Claim Your 250,000 Free Chips' with a blue 'OPEN' button. Below this, a news article is displayed with the headline 'Cinco millones de afiliados retiraron S/ 32,100 millones de sus fondos de AFP en pandemia'. The article text mentions that 1.9 million people withdrew funds during the health crisis, totaling 539,000 people. A small image of hands holding money is visible. The right side of the browser shows the 'Elements' panel with the DOM tree. The selected element is a link with the href </tu-dinero/cinco-millones-de-afiliados-retiraron-s-32100-millones-de-sus-fondos-de-afp-en-pandemia-noticia/>. The 'Console' panel at the bottom shows 'What's New' and 'Highlights from the Chrome 88 update'.

Claim Your 250,000 Free Chips

Play free poker online 24/7 with the official World Series of Poker game

PlayWSOP

OPEN

a.story-item__title.block.overflow-hidden.primary-font.line-h-xs.mt-10 375 x 84

Cinco millones de afiliados retiraron S/ 32,100 millones de sus fondos de AFP en pandemia

De los que hicieron retiros en la crisis sanitaria, 1,9 millones se quedaron con S/ 0 en sus cuentas en AFP. De estos, 539,000 son mayores de 51 ...

Elements Console Sources Network Performance Memory

```
large pb-10">AFP</h1>
<div id="fusion-static-exit:0fc6iqD5xrU" style="display: none;"></div>
<div id="fusion-static-enter:0fu6YyD5xrr" style="display: none;"></div>
<div>
  <div class="story-item w-full pr-20 pl-20 pb-20 mb-20 border-b-1 border-solid border-gray md:pl-0 md:pr-0 lg:p-0">
    <div class="story-item__bottom flex lg:pb-15">
      <div class="story-item__left flex flex-col justify-between pr-20">
        <div class="story-item__top flex items-center md:flex-col md:items-start">...</div>
        <div class="story-item__information-box w-full">
          <h2 itemprop="name" class="story-item__content-title overflow-hidden">
            <a itemprop="url" class="story-item__title block overflow-hidden primary-font line-h-xs mt-10" href="/tu-dinero/cinco-millones-de-afiliados-retiraron-s-32100-millones-de-sus-fondos-de-afp-en-pandemia-noticia/">...</a> == $0
          </h2>
            <p itemprop="description" class="story-item__subtitle overflow-hidden hidden mt-10 mb-10 text-md text-gray-200 line-h-xs">...</p>
            <a itemprop="url" href="/tu-dinero/" class="story-item__section story-item__section--desktop text-sm text-black md:mb-15 hidden">Tu Dinero</a>
          </div>
            <div class="story-item__author-wrapper hidden">...</div>
          </div>
            <figure class="story-item__right position-relative overflow-hidden">...</figure>
          </div>
        </div>
      </div>
    </div>
  </div>
</div>
```

... .content-title.overflow-hidden a.story-item__title.block.overflow-hidden.primary-font.line-h-xs.mt-10 ...

Console What's New

Highlights from the Chrome 88 update

Método 4

Web Scraping Dinámico
(*RSelenium*)

Web Scraping: Página Web Dinámica

Inspeccionar Página Web (I)

https://www.sbs.gob.pe/app/pu/CCID/Paginas/vp_rentafija.aspx

← → ↻ sbs.gob.pe/app/pu/CCID/Paginas/vp_rentafija.aspx

SUPERINTENDENCIA DE BANCA, SEGUROS Y AFP
República del Perú

RENTA FIJA

Fecha de Proceso: 12/02/2021 Tipo de Moneda: Todas Emisor: Todos Rating Emisión: Todos **Consultar**

Exportar a excel

Nemónico	ISIN/Identif.	Emisor	Moneda	P. Limpio (%)	TIR %	Origen(*)
CHAV1BC1A	PEP20950M010	A&G Chavin	USD	100.026	9.004	
CHAV1BC1B	PEP20950M028	A&G Chavin	USD	99.995	9.022	
CHAV1BC1C	PEP20950M036	A&G Chavin	USD	99.981	9.052	
CHAV1BC1D	PEP20950M044	A&G Chavin	USD	99.929	9.075	
CHAV1BC1E	PEP20950M051	A&G Chavin	USD	99.988	9.054	
CHAV1BC1F	PEP20950M069	A&G Chavin	USD	100.014	9.043	
CHAV1CP1I	PEP20950V417	A&G Chavin	USD	99.642	7.891	
CHAV1CP1J	PEP20950V425	A&G Chavin	USD	97.556	7.982	
CHAV1CP1K	PEP20950V433	A&G Chavin	USD	96.817	8.016	
CHAV1CP1L	PEP20950V441	A&G Chavin	USD	96.374	8.038	
CHAV1CP1M	PEP20950V458	A&G Chavin	USD	94.630	8.137	
ALIC1DBC1A	PEP21400M106	ALICORP S.A.	PEN	115.192	2.172	
ALIC1DBC2A	PEP21400M064	ALICORP S.A.	PEN	120.964	3.428	
ALIC1DBC4A	PEP21400M080	ALICORP S.A.	PEN	112.808	1.584	
ATRIA1BC1A	PEP70002M019	Atria Energía	USD	101.107	5.057	
ATRIA1BC1B	PEP70002M027	Atria Energía	USD	100.864	5.092	
BRIP14BC3A	PEP14300M198	BCO RIPLEY	PEN	100.460	4.186	
BRIP14BC3B	PEP14300M206	BCO RIPLEY	PEN	101.161	4.465	
BRIP14BC4A	PEP14300M214	BCO RIPLEY	PEN	100.765	4.215	
BRIP14BC4B	PEP14300M222	BCO RIPLEY	PEN	100.160	4.462	
BRIP15BC1A	PEP14300M230	BCO RIPLEY	PEN	99.216	4.570	
CD02MAR21	CD02MAR21	BCRP	PEN	99.987	.261	
CD02MAR23	CD02MAR23	BCRP	PEN	98.044	.955	
CD03AGO21	CD03AGO21	BCRP	PEN	99.869	.274	
CD03ENE23	CD03ENE23	BCRP	PEN	98.383	.854	
CD03JUN21	CD03JUN21	BCRP	PEN	99.921	.258	
CD03MAR22	CD03MAR22	BCRP	PEN	99.545	.428	
CD04ENE22	CD04ENE22	BCRP	PEN	99.664	.373	
CD04MAY21	CD04MAY21	BCRP	PEN	99.938	.275	
CD06ABR21	CD06ABR21	BCRP	PEN	99.962	.255	

Elements Console Sources Network

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head id="Head1">...</head>
  <body>
    <form name="form1" method="post" action="vp_rentafija.aspx">
      <div>...</div>
      <script type="text/javascript">...</script>
      <script src="/app/pu/CCID/WebResource.axd?d=F668u2TtttDvdcY00w06Dd1Lxhthh80aay0bUwE8dw">...</script>
    </form>
  </body>
</html>
```

Atrás Alt+Flecha izquierda
Reenviar Alt+Flecha derecha
Volver a cargar Ctrl+R
Guardar como... Ctrl+S
Imprimir... Ctrl+P
Transmitir...
Crear un código QR para esta página
Traducir a español
Ver código fuente de página Ctrl+U
Inspeccionar Ctrl+Shift+I

Web Scraping: Página Web Dinámica

Inspeccionar Página Web (II)

The screenshot shows a web application interface on the left and its HTML structure in the browser's developer tools on the right. A red arrow points from the 'Consultar' button in the web interface to the corresponding HTML code in the developer tools.

Web Interface:

Moneda: Emisor: Rating Emisión:

Moneda	P. Limpio (%)	TIR %	Origen(*)	Spreads
USD	100.026	9.004		8.102
USD	99.995	9.022		8.115
USD	99.981	9.052		8.138
USD	99.929	9.075		8.155
USD	99.988	9.054		8.122
USD	100.014	9.043		8.106
USD	99.642	7.891		7.489
USD	97.556	7.982		7.549
USD	96.817	8.016		7.572
USD	96.374	8.038		7.586
USD	94.630	8.137		7.646
PEN	115.192	2.172		.587
PEN	120.964	3.428		.690
PEN	112.808	1.584		.662
USD	101.107	5.057		4.145
USD	100.864	5.092		4.146
PEN	100.460	4.186		3.850
PEN	101.161	4.465		3.926
PEN	100.765	4.215		3.858
PEN	100.160	4.462		3.928
PEN	99.216	4.570		3.864
PEN	99.987	.261		.000
PEN	98.044	.955		.000
PEN	99.869	.274		.000
PEN	98.383	.854		.000
PEN	99.921	.258		.000

HTML Structure (Developer Tools):

```
<div>...</div>
<div>
  <table width="100%" border="0" align="center" cellpadding="0" cellspacing="0"
    class="Apli_back3">...</table>
  <div id="divformulario">
    <br>
    <div class="APII_titulo" style="text-align:left;">...</div>
    <table cellpadding="0" border="0" class="APII_conteTabla2" width="100%"
      style="width:2350px">
      <tbody>
        <tr>
          <td>
            <table cellpadding="5">
              <tbody>
                <tr>...</tr>
                <tr>...</tr>
                <tr>
                  <td colspan="2">
                    <td colspan="8" align="right">
                      <input type="submit" name="btnConsultar" value="Consultar" id=
                        "btnConsultar" class="APII_boton"> == $0
                    </td>
                  </tr>
                </tbody>
              </table>
            </td>
          </tr>
        </tbody>
      </table>
    </div>
  </div>
</div>
```

Web Scraping: Página Web Dinámica

RSelenium: Abrir Conexión

Conexión con servidor

```
server <- phantomjs(port = 5004L)  
browser <- remoteDriver(browserName = "phantomjs", port = 5004L)
```

Abrir conexión

```
browser$open()
```

Navegar en una página web

```
browser$navigate(URL)
```

Mostrar navegador en el Viewer

```
browser$screenshot(display = TRUE)
```



Nemónico	ISIN/Identif.	Emisor	Moneda	P. Limpio (%)
ALIC1DBC1A	PEP21400M106	ALICORP S.A.	PEN	108.524
ALIC1DCP1A	PEP21400VA24	ALICORP S.A.	PEN	98.592
ALICO3BC2A	PEP21400M064	ALICORP S.A.	PEN	112.022
ALICO3BC4A	PEP21400M080	ALICORP S.A.	PEN	113.001
BFAL1BC4A	PEP14150M056	BANCO FALABELLA	PEN	100.794
BFAL1BC7A	PEP14150M072	BANCO FALABELLA	PEN	100.482
BFAL1BC1A	PEP14150M188	BANCO FALABELLA	PEN	100.088

Web Scraping: Página Web Dinámica

RSelenium: Métodos para extraer un elemento web en R

- **id**

```
$findElement(using = 'id', value = "btnConsultar")
```

- **class**

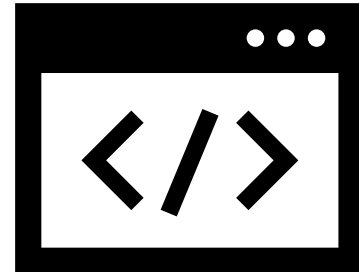
```
$findElement(using = 'class', value = " APLI_boton")
```

- **css selector**

```
$findElement(using = 'css select', value = "#btnConsultar")
```

- **xpath**

```
$findElement(using = 'xpath', value = "//*[@id='btnConsultar']")
```



Web Scraping: Página Web Dinámica

RSelenium: Acciones y Almacenamiento de Información

Dar click al elemento

```
$clickElement()
```

Extraer el documento HTML

```
browser$getPageSource[[1]]
```

Colocar en un formato leíble en R

```
htmlParse()
```

Extraer datos de las tablas HTML

```
readHTMLTable()
```

