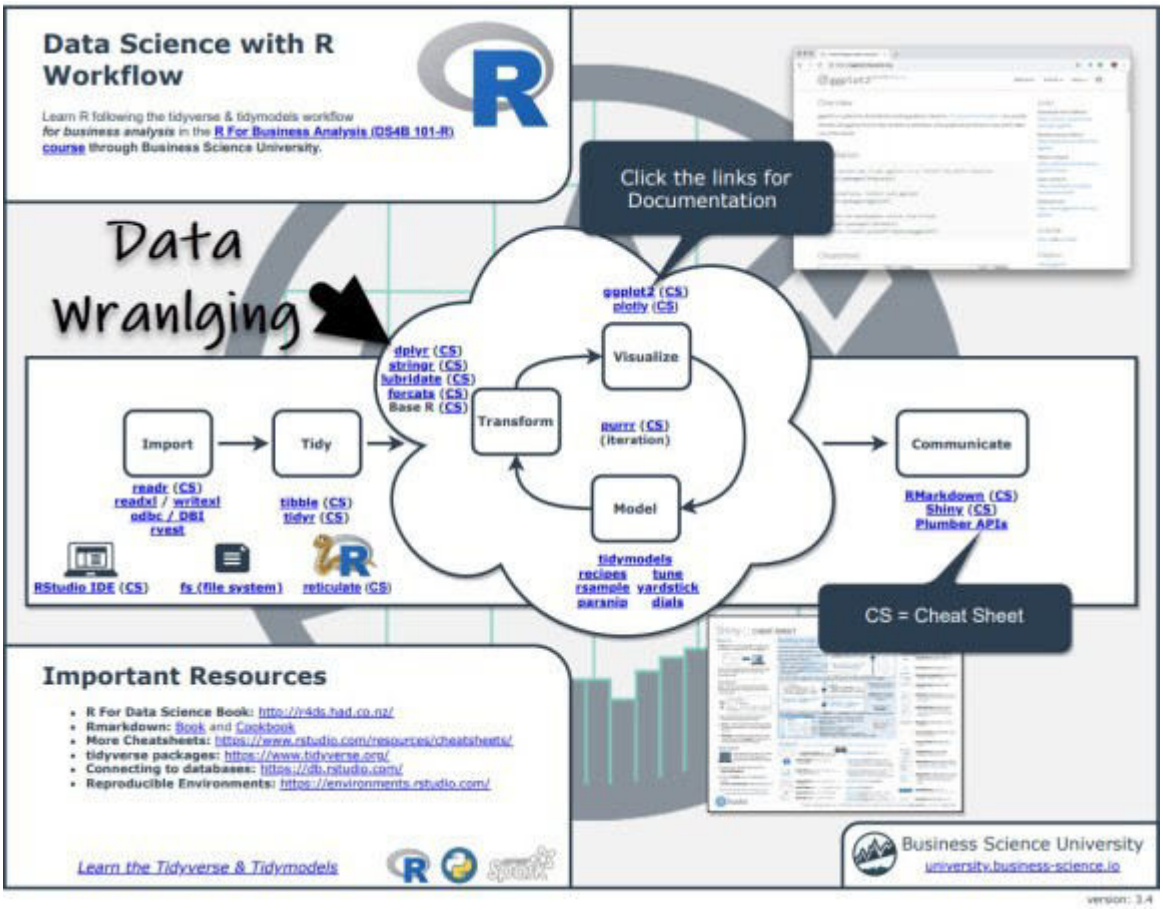
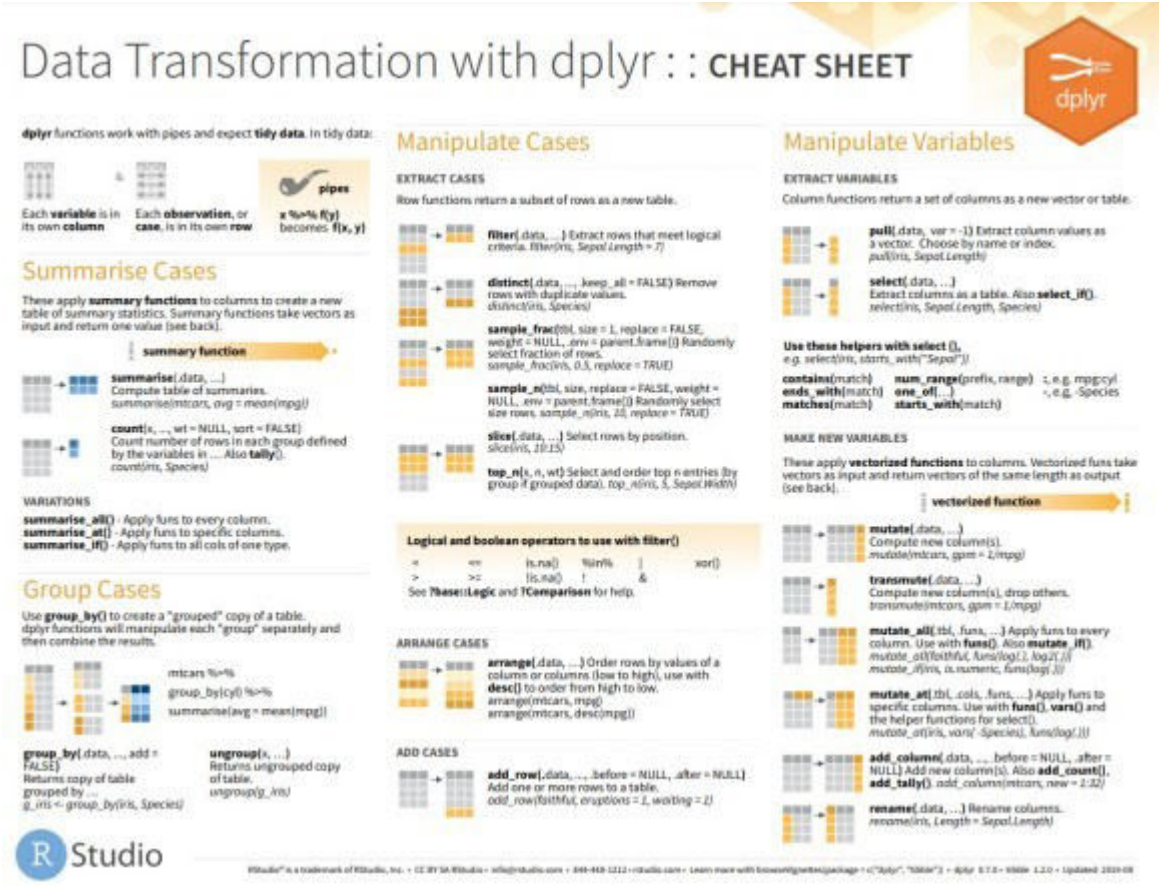
Before we get started, get the Cheat Sheet

`Datapasta` is great for importing raw data from HTML tables on the web. But, you'll still need to learn how to transform / wrangle the data and produce visualizations. For those topics, I'll use the Ultimate R Cheat Sheet to refer to `dplyr` and `ggplot2` code in my workflow.

Quick example – Clicking the "CS" next to "dplyr" opens the Data Transformation with Dplyr Cheat Sheet.

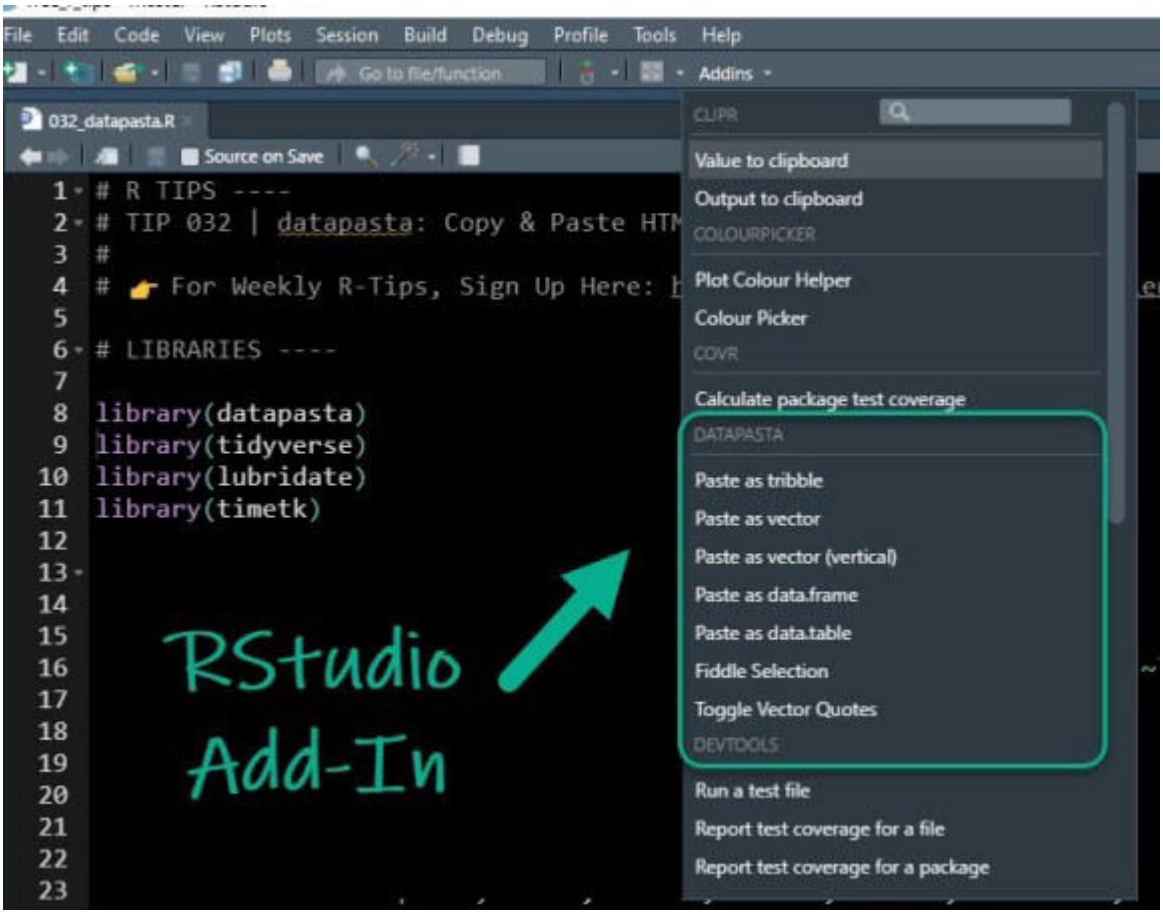


Now you're ready to quickly reference `dplyr` functions. Ok, onto the tutorial.



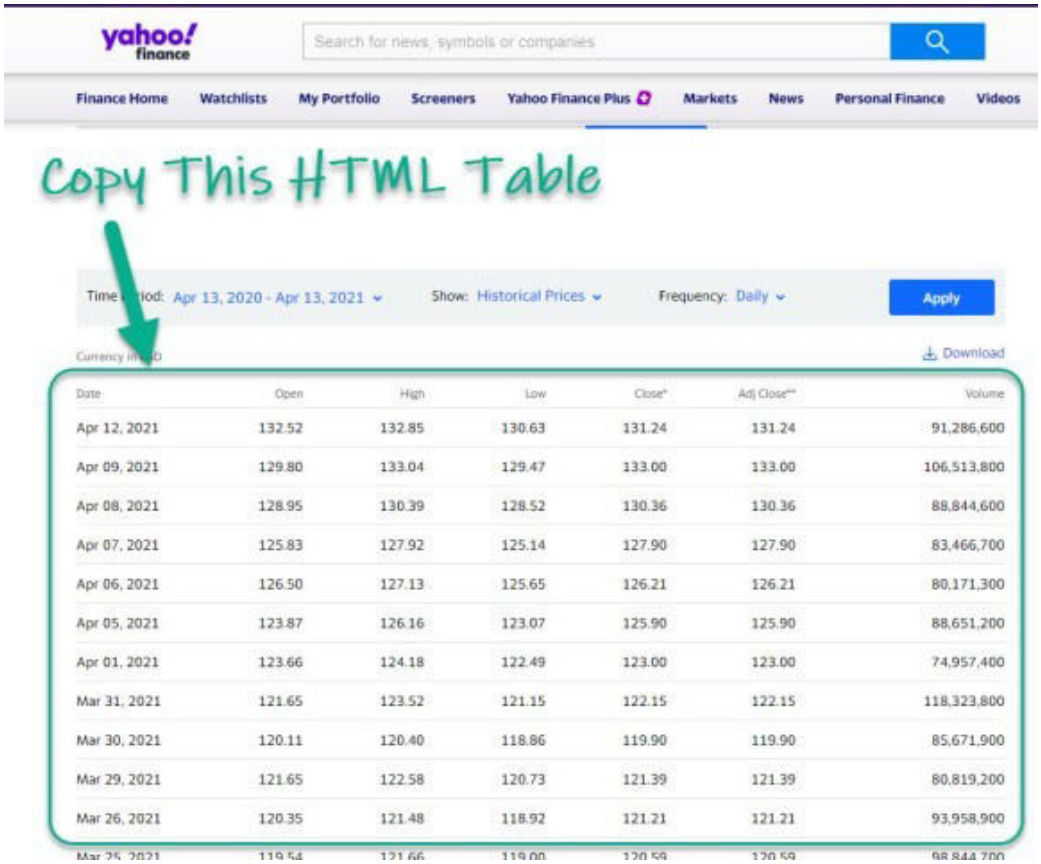
Datapasta RStudio Addin

Datapasta contains an RStudio Add-In for Pasting web-tables stored in your “clipboard” (what happens when you “copy” something).



Example 1: Copying Stock Data from Yahoo! Finance

Let's go through a quick example. We can navigate to Yahoo! Finance and search for a ticker symbol like AAPL.

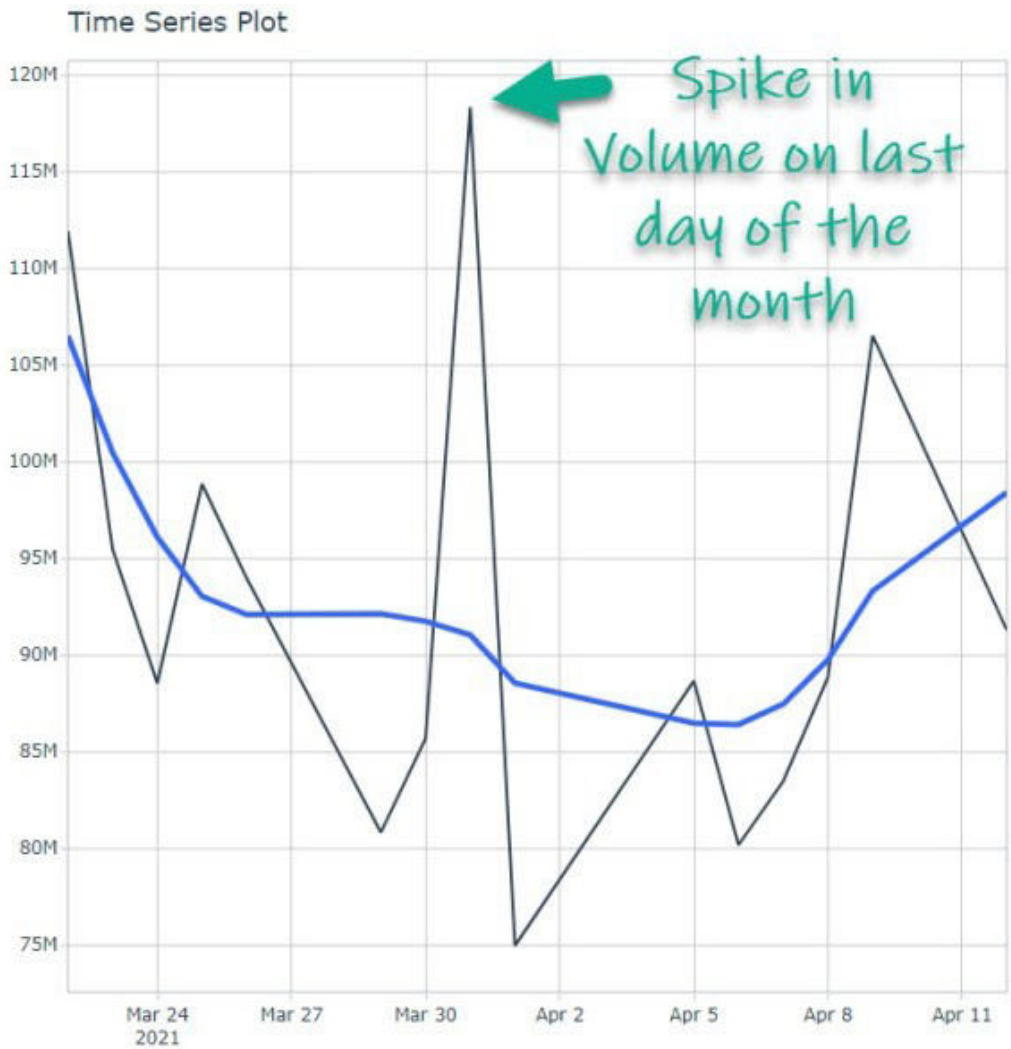


Source: Yahoo! Finance

Next, use the Datapasta Addin to **"paste as tribble"**. This pastes our data into our R script file.

```
tibble::tribble(
            ~Date,  ~Open,  ~High,   ~Low, ~`Close*`, ~`Adj.Close**`,     ~Volume,
   "Apr 12, 2021", 132.52, 132.85, 130.63,    131.24,         131.24,    91286600,
   "Apr 09, 2021",  129.8, 133.04, 129.47,       133,            133,   106513800,
   "Apr 08, 2021", 128.95, 130.39, 128.52,    130.36,         130.36,    88844600,
   "Apr 07, 2021", 125.83, 127.92, 125.14,     127.9,          127.9,    83466700,
   "Apr 06, 2021",  126.5, 127.13, 125.65,    126.21,         126.21,    80171300,
   "Apr 05, 2021", 123.87, 126.16, 123.07,     125.9,          125.9,    88651200,
   "Apr 01, 2021", 123.66, 124.18, 122.49,       123,            123,    74957400,
   "Mar 31, 2021", 121.65, 123.52, 121.15,    122.15,         122.15,   118323800,
   "Mar 30, 2021", 120.11,  120.4, 118.86,     119.9,          119.9,    85671900,
   "Mar 29, 2021", 121.65, 122.58, 120.73,    121.39,         121.39,    80819200,
   "Mar 26, 2021", 120.35, 121.48, 118.92,    121.21,         121.21,    93958900,
   "Mar 25, 2021", 119.54, 121.66,    119,    120.59,         120.59,    98844700,
   "Mar 24, 2021", 122.82,  122.9, 120.07,    120.09,         120.09,    88530500,
   "Mar 23, 2021", 123.33, 124.24, 122.14,    122.54,         122.54,    95467100,
   "Mar 22, 2021", 120.33, 123.87, 120.26,    123.39,         123.39,   111912300
   )
```

Next, use `dplyr` and `timetk` to wrangle and visualize the data. (Refer to the ultimate R cheat sheet for documentation on dplyr and timetk). We can see a spike in volume on last day of the month.



Code available in our Free R-Tips Github Repository

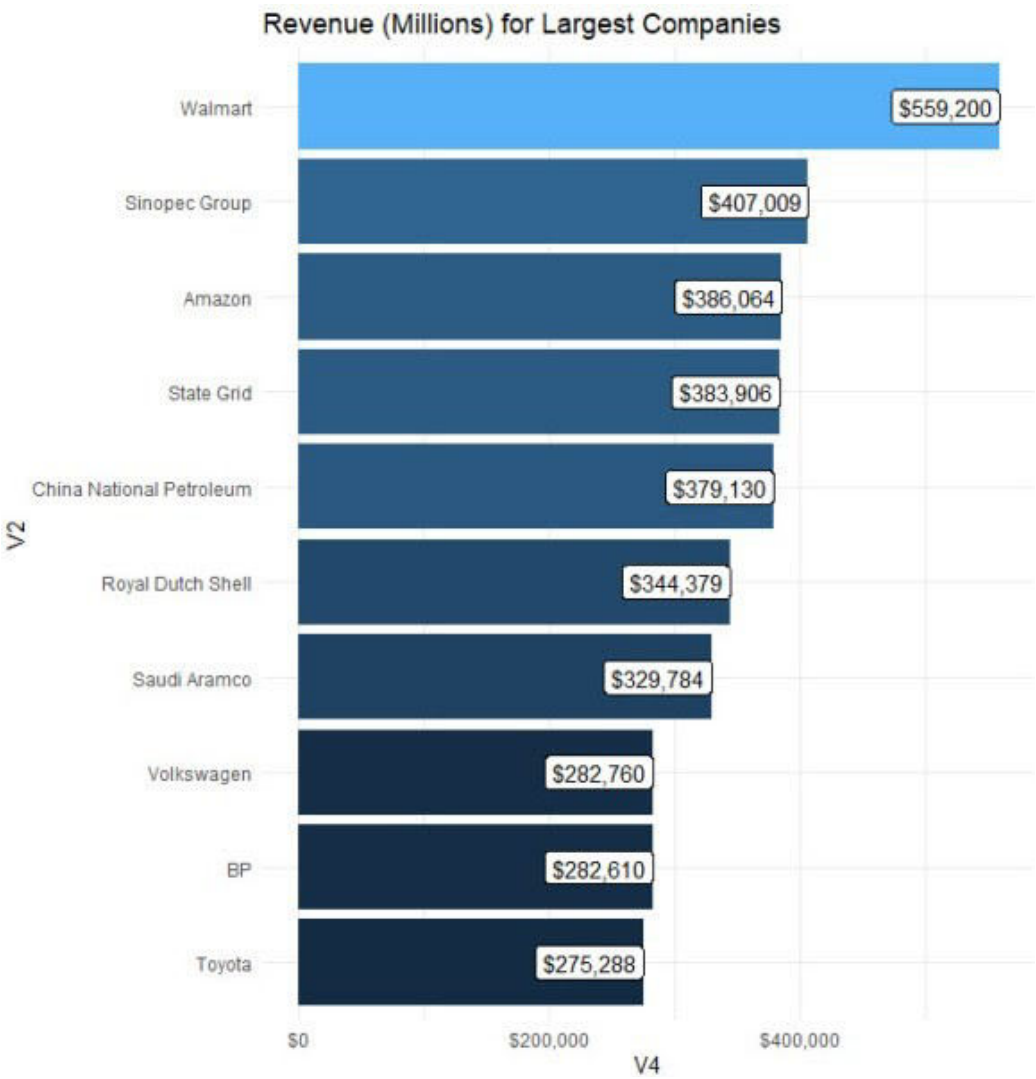
Example 2: Getting Revenue Data for World Largest Companies From Wikipedia

First, head over to Wikipedia and search for the “list of largest companies”.

Rank	Name	Industry	Revenue (USD Millions)	Profit (USD Millions)	Employees	Headquarters	Ref
1	Walmart	Retail	▲ $559,200	$19,742	2,200,000	United States	[4]
2	Sinopec Group	Oil and gas	▼ $407,009	$6,793	582,648	China	[5]
3	Amazon	Retail	▲ $386,064	$17,377	1,125,300	United States	[6]
4	State Grid	Electricity	▼ $383,906	$7,970	907,677	China	[7]
5	China National Petroleum	Oil and gas	▼ $379,130	$4,433	1,344,410	China	[8]
6	Royal Dutch Shell	Oil and gas	▼ $344,379	$15,842	83,000	Netherlands	[9]
7	Saudi Aramco	Oil and gas	▼ $329,784	$88,211	79,000	Saudi Arabia	[10]
8	Volkswagen	Automotive	▲ $282,760	$15,542	671,205	Germany	[11]
9	BP	Oil and gas	▼ $282,610	$4,026	72,500	United Kingdom	[12]
10	Toyota	Automotive	▲ $275,288	$19,096	359,542	Japan	[13]
11	Apple	Electronics	▼ $274,515	$55,256	137,000	United States	[14]
12	ExxonMobil	Oil and gas	▼ $264,938	$14,340	74,900	United States	[15]
13	CVS Health	Healthcare	▲ $256,776	$6,634	290,000	United States	[16]
14	Berkshire Hathaway	Financials	▲ $254,616	$81,417	391,500	United States	[17]

Source: Wikipedia

Use **datapasta** to “paste as data.table”. Then do some data wrangling with **dplyr**. Then visualize with **ggplot2**. And in a few lines of code you can create this chart showing that Walmart is dominating in Revenue. (Refer to the ultimate R cheat sheet for documentation on dplyr and ggplot2).



Code available in our Free R-Tips Github Repository

In Summary

You just quickly scraped HTML tables using the copy-and-paster Rstudio Add-In known as **datapasta** . This is an amazing productivity boost!!