Data Journalism

Paul Ryan speeches using words over time

Mock data

**Advanced scrapping module???**

**## Create coding cells as needed**

**## assign variables to data**

**latino\_jaywalking\_23 = 154**

**black\_jaywalking\_23 = 276**

**total\_jaywalking\_23 = 467**

**##write dynamic sentence**

**Week9**

**Digital PDF**

**You copy the text and paste it but it wouldn’t keep the format when pasted into excel.**

**Another technique**

**Tabula App that pulls tables out of a pdf**

**Tabula GUI is a Python library, method, …**

**These packages allow journalists to have superpowers**

**Concat needs a list**

**Ignore index = true**

**Markdown cheatsheet**

**Basic syntax**

**“w” stands for write**

**Parameter place holder**

**Arguments specific**

**HTML- Static URL**

**HTML MIX- Cleaning**

**JAVA SCRIPT Scrape- find**

**JSON- Dynamic (apart of java script)**

**Data is stored in the server not in the URL**

**You have to go get it**

**Nested- dictionaries within dictionaries- flattening of lists**

**API- Gov data, Sensitive Data (big data)- pass key to access**

**Soup- turning it into a language python can understand**

**Parser- teaching python how to read tags, this is an html file**

**HTML parsing is the process of analyzing a string of HTML code to identify its structure and extract relevant information. This involves breaking down the HTML into its constituent elements such as tags, attributes, and text content.**