DATA607 Lab 7

Samuel C

2025-03-16

Overview

In this lab, we have been tasked with creating many different file types from a list of inventory at a local store. These file types include JSON, HTML, XML and Parquet. Once the data is transformed into the correct format, I will export it to the file type for easy storage. Additionally I will go over the pros and cons of each file type.

Load Packages

In this lab, I researched many different packages to help format data appropriately for export.

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ---------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4      v readr     2.1.5
## v forcats   1.0.0      v stringr   1.5.1
## v ggplot2   3.5.1      v tibble    3.2.1
## v lubridate 1.9.4      v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(arrow)
```

```
## Warning: package 'arrow' was built under R version 4.4.3
```

```
##
## Attaching package: 'arrow'
##
## The following object is masked from 'package:lubridate':
##
##     duration
##
## The following object is masked from 'package:utils':
##
##     timestamp
```

```r
library(repurrrsive)
```

```
## Warning: package 'repurrrsive' was built under R version 4.4.3
```

```r
library(jsonlite)
```

```
##
## Attaching package: 'jsonlite'
##
## The following object is masked from 'package:purrr':
##
##     flatten
```

```r
library(XML)
library(xml2)
library(tableHTML)
```

```
## Warning: package 'tableHTML' was built under R version 4.4.3
```

Once these were installed I could begin.

Import and Cleaning

I first saved the data given in a pdf into a CSV file, which I then saved to Github to import.

```r
lab7data <- read.csv("https://raw.githubusercontent.com/scrummett/DATA607/refs/heads/main/lab7data.csv"
```

However, due to the number of columns and the number of values separated by commas, I must adjust the table.

```r
colnames(lab7data) <- c(colnames(lab7data)[-1], "Variation.Details2")
lab7data <- lab7data |>
  mutate(Variation.Details = paste(Variation.Details, Variation.Details2, sep = ",")) |>
  select(!Variation.Details2)
```

Here I have shifted the column names left by one, named the furthest right column, then combined the values from the two columns containing item details into one column. From here we can create our different files.

JSON

I will start with our easiest files to code, that being a JSON file. Fortunately, the package "jsonlite" makes this incredibly easy for us.

```r
toJSON(lab7data)
```

```
## [{"Category":" Electronics","Item.Name":"Smartphone","Item.ID":101,"Brand":"TechBrand","Price":699.9!
```

This converts our data into JSON format, from which we can export it.

```r
exportJSON <- toJSON(lab7data)
write(exportJSON, "C:/Users/crumm/OneDrive/Documents/MSDS Spring 2025/DATA607 Labs/lab7data.JSON")
JSONtest <- fromJSON("C:/Users/crumm/OneDrive/Documents/MSDS Spring 2025/DATA607 Labs/lab7data.JSON")
print(JSONtest)
```

```
##              Category      Item.Name Item.ID      Brand   Price Variation.ID
## 1        Electronics     Smartphone     101  TechBrand  699.99        101-A
## 2        Electronics     Smartphone     101  TechBrand  699.99        101-B
## 3        Electronics         Laptop     102 CompuBrand 1099.99        102-A
## 4        Electronics         Laptop     102 CompuBrand 1099.99        102-B
## 5    Home Appliances   Refrigerator     201   HomeCool  899.99        201-A
## 6    Home Appliances   Refrigerator     201   HomeCool  899.99        201-B
## 7    Home Appliances Washing Machine     202  CleanTech  499.99        202-A
## 8    Home Appliances Washing Machine     202  CleanTech  499.99        202-B
## 9           Clothing        T-Shirt     301  FashionCo   19.99        301-A
## 10          Clothing        T-Shirt     301  FashionCo   19.99        301-B
## 11          Clothing        T-Shirt     301  FashionCo   19.99        301-C
## 12          Clothing          Jeans     302 DenimWorks   49.99        302-A
## 13          Clothing          Jeans     302 DenimWorks   49.99        302-B
## 14             Books   Fiction Novel     401          -   14.99        401-A
## 15             Books   Fiction Novel     401          -   14.99        401-B
## 16             Books Non-Fiction Guide 402          -   24.99        402-A
## 17             Books Non-Fiction Guide 402          -   24.99        402-B
## 18  Sports Equipment     Basketball     501 SportsGear   29.99        501-A
## 19  Sports Equipment   Tennis Racket     502  RacketPro   89.99        502-A
## 20  Sports Equipment   Tennis Racket     502  RacketPro   89.99        502-B
##                          Variation.Details
## 1             Color: Black, Storage: 64GB
## 2            Color: White, Storage: 128GB
## 3           Color: Silver, Storage: 256GB
## 4       Color: Space Gray, Storage: 512GB
## 5   Color: Stainless Steel, Capacity: 20 cu ft
## 6           Color: White, Capacity: 18 cu ft
## 7        Type: Front Load, Capacity: 4.5 cu ft
## 8         Type: Top Load, Capacity: 5.0 cu ft
## 9                      Color: Blue, Size: S
## 10                      Color: Red, Size: M
## 11                    Color: Green, Size: L
## 12                 Color: Dark Blue, Size: 32
## 13                Color: Light Blue, Size: 34
## 14        Format: Hardcover, Language: English
## 15        Format: Paperback, Language: Spanish
## 16            Format: eBook, Language: English
## 17         Format: Paperback, Language: French
## 18                Size: Size 7, Color: Orange
## 19            Material: Graphite, Color: Black
## 20            Material: Aluminum, Color: Silver
```

Here we have saved it as a JSON file, as well as called that same file to test it to see if our save has worked. This table demonstrates that we have saved it correctly.

Parquet

Next is saving our data to a parquet file, which is made simple by the package "arrow".

```
write_parquet(lab7data, "C:/Users/crumm/OneDrive/Documents/MSDS Spring 2025/DATA607 Labs/lab7data.parque
```

Here we have simply taken our data and saved it as a parquet file.

```
lab7parquet <- open_dataset("C:/Users/crumm/OneDrive/Documents/MSDS Spring 2025/DATA607 Labs/lab7data.pa
head(lab7parquet) |> collect()
```

```
## # A tibble: 6 x 7
##    Category         Item.Name Item.ID Brand Price Variation.ID Variation.Details
##    <chr>            <chr>        <int> <chr> <dbl> <chr>        <chr>
## 1 " Electronics"    Smartpho~      101 Tech~  700. 101-A        Color: Black, St~
## 2 " Electronics"    Smartpho~      101 Tech~  700. 101-B        Color: White, St~
## 3 " Electronics"    Laptop         102 Comp~ 1100. 102-A        Color: Silver, S~
## 4 " Electronics"    Laptop         102 Comp~ 1100. 102-B        Color: Space Gra~
## 5 " Home Appliance~ Refriger~      201 Home~  900. 201-A        Color: Stainless~
## 6 " Home Appliance~ Refriger~      201 Home~  900. 201-B        Color: White, Ca~
```

Again we have called that saved file and presented it as a table, showing that we have successfully created a parquet file.

HTML

Again, this is made simple ny a package, this package being called "tableHTML". It takes data from a table and converts it into a table in html.

```
lab7HTML <- tableHTML(lab7data)
write_tableHTML(lab7HTML, file = "C:/Users/crumm/OneDrive/Documents/MSDS Spring 2025/DATA607 Labs/lab7da
print(lab7HTML)
```

```
##
## <table style="border-collapse:collapse;" class=table_4353 border=1>
## <thead>
## <tr>
##   <th id="tableHTML_header_1"> </th>
##   <th id="tableHTML_header_2">Category</th>
##   <th id="tableHTML_header_3">Item.Name</th>
##   <th id="tableHTML_header_4">Item.ID</th>
##   <th id="tableHTML_header_5">Brand</th>
##   <th id="tableHTML_header_6">Price</th>
##   <th id="tableHTML_header_7">Variation.ID</th>
##   <th id="tableHTML_header_8">Variation.Details</th>
## </tr>
## </thead>
## <tbody>
## <tr>
##   <td id="tableHTML_rownames">1</td>
##   <td id="tableHTML_column_1"> Electronics</td>
##   <td id="tableHTML_column_2">Smartphone</td>
##   <td id="tableHTML_column_3">101</td>
##   <td id="tableHTML_column_4">TechBrand</td>
##   <td id="tableHTML_column_5">699.99</td>
##   <td id="tableHTML_column_6">101-A</td>
##   <td id="tableHTML_column_7">Color: Black, Storage: 64GB</td>
```

```
## </tr>
## <tr>
##   <td id="tableHTML_rownames">2</td>
##   <td id="tableHTML_column_1"> Electronics</td>
##   <td id="tableHTML_column_2">Smartphone</td>
##   <td id="tableHTML_column_3">101</td>
##   <td id="tableHTML_column_4">TechBrand</td>
##   <td id="tableHTML_column_5">699.99</td>
##   <td id="tableHTML_column_6">101-B</td>
##   <td id="tableHTML_column_7">Color: White, Storage: 128GB</td>
## </tr>
## <tr>
##   <td id="tableHTML_rownames">3</td>
##   <td id="tableHTML_column_1"> Electronics</td>
##   <td id="tableHTML_column_2">Laptop</td>
##   <td id="tableHTML_column_3">102</td>
##   <td id="tableHTML_column_4">CompuBrand</td>
##   <td id="tableHTML_column_5">1099.99</td>
##   <td id="tableHTML_column_6">102-A</td>
##   <td id="tableHTML_column_7">Color: Silver, Storage: 256GB</td>
## </tr>
## <tr>
##   <td id="tableHTML_rownames">4</td>
##   <td id="tableHTML_column_1"> Electronics</td>
##   <td id="tableHTML_column_2">Laptop</td>
##   <td id="tableHTML_column_3">102</td>
##   <td id="tableHTML_column_4">CompuBrand</td>
##   <td id="tableHTML_column_5">1099.99</td>
##   <td id="tableHTML_column_6">102-B</td>
##   <td id="tableHTML_column_7">Color: Space Gray, Storage: 512GB</td>
## </tr>
## <tr>
##   <td id="tableHTML_rownames">5</td>
##   <td id="tableHTML_column_1"> Home Appliances</td>
##   <td id="tableHTML_column_2">Refrigerator</td>
##   <td id="tableHTML_column_3">201</td>
##   <td id="tableHTML_column_4">HomeCool</td>
##   <td id="tableHTML_column_5">899.99</td>
##   <td id="tableHTML_column_6">201-A</td>
##   <td id="tableHTML_column_7">Color: Stainless Steel, Capacity: 20 cu ft</td>
## </tr>
## <tr>
##   <td id="tableHTML_rownames">6</td>
##   <td id="tableHTML_column_1"> Home Appliances</td>
##   <td id="tableHTML_column_2">Refrigerator</td>
##   <td id="tableHTML_column_3">201</td>
##   <td id="tableHTML_column_4">HomeCool</td>
##   <td id="tableHTML_column_5">899.99</td>
##   <td id="tableHTML_column_6">201-B</td>
##   <td id="tableHTML_column_7">Color: White, Capacity: 18 cu ft</td>
## </tr>
## <tr>
##   <td id="tableHTML_rownames">7</td>
##   <td id="tableHTML_column_1"> Home Appliances</td>
```

```
##    <td id="tableHTML_column_2">Washing Machine</td>
##    <td id="tableHTML_column_3">202</td>
##    <td id="tableHTML_column_4">CleanTech</td>
##    <td id="tableHTML_column_5">499.99</td>
##    <td id="tableHTML_column_6">202-A</td>
##    <td id="tableHTML_column_7">Type: Front Load, Capacity: 4.5 cu ft</td>
## </tr>
## <tr>
##    <td id="tableHTML_rownames">8</td>
##    <td id="tableHTML_column_1"> Home Appliances</td>
##    <td id="tableHTML_column_2">Washing Machine</td>
##    <td id="tableHTML_column_3">202</td>
##    <td id="tableHTML_column_4">CleanTech</td>
##    <td id="tableHTML_column_5">499.99</td>
##    <td id="tableHTML_column_6">202-B</td>
##    <td id="tableHTML_column_7">Type: Top Load, Capacity: 5.0 cu ft</td>
## </tr>
## <tr>
##    <td id="tableHTML_rownames">9</td>
##    <td id="tableHTML_column_1"> Clothing</td>
##    <td id="tableHTML_column_2">T-Shirt</td>
##    <td id="tableHTML_column_3">301</td>
##    <td id="tableHTML_column_4">FashionCo</td>
##    <td id="tableHTML_column_5">19.99</td>
##    <td id="tableHTML_column_6">301-A</td>
##    <td id="tableHTML_column_7">Color: Blue, Size: S</td>
## </tr>
## <tr>
##    <td id="tableHTML_rownames">10</td>
##    <td id="tableHTML_column_1"> Clothing</td>
##    <td id="tableHTML_column_2">T-Shirt</td>
##    <td id="tableHTML_column_3">301</td>
##    <td id="tableHTML_column_4">FashionCo</td>
##    <td id="tableHTML_column_5">19.99</td>
##    <td id="tableHTML_column_6">301-B</td>
##    <td id="tableHTML_column_7">Color: Red, Size: M</td>
## </tr>
## <tr>
##    <td id="tableHTML_rownames">11</td>
##    <td id="tableHTML_column_1"> Clothing</td>
##    <td id="tableHTML_column_2">T-Shirt</td>
##    <td id="tableHTML_column_3">301</td>
##    <td id="tableHTML_column_4">FashionCo</td>
##    <td id="tableHTML_column_5">19.99</td>
##    <td id="tableHTML_column_6">301-C</td>
##    <td id="tableHTML_column_7">Color: Green, Size: L</td>
## </tr>
## <tr>
##    <td id="tableHTML_rownames">12</td>
##    <td id="tableHTML_column_1"> Clothing</td>
##    <td id="tableHTML_column_2">Jeans</td>
##    <td id="tableHTML_column_3">302</td>
##    <td id="tableHTML_column_4">DenimWorks</td>
##    <td id="tableHTML_column_5">49.99</td>
```

```
##     <td id="tableHTML_column_6">302-A</td>
##     <td id="tableHTML_column_7">Color: Dark Blue, Size: 32</td>
##   </tr>
##   <tr>
##     <td id="tableHTML_rownames">13</td>
##     <td id="tableHTML_column_1"> Clothing</td>
##     <td id="tableHTML_column_2">Jeans</td>
##     <td id="tableHTML_column_3">302</td>
##     <td id="tableHTML_column_4">DenimWorks</td>
##     <td id="tableHTML_column_5">49.99</td>
##     <td id="tableHTML_column_6">302-B</td>
##     <td id="tableHTML_column_7">Color: Light Blue, Size: 34</td>
##   </tr>
##   <tr>
##     <td id="tableHTML_rownames">14</td>
##     <td id="tableHTML_column_1"> Books</td>
##     <td id="tableHTML_column_2">Fiction Novel</td>
##     <td id="tableHTML_column_3">401</td>
##     <td id="tableHTML_column_4">-</td>
##     <td id="tableHTML_column_5">14.99</td>
##     <td id="tableHTML_column_6">401-A</td>
##     <td id="tableHTML_column_7">Format: Hardcover, Language: English</td>
##   </tr>
##   <tr>
##     <td id="tableHTML_rownames">15</td>
##     <td id="tableHTML_column_1"> Books</td>
##     <td id="tableHTML_column_2">Fiction Novel</td>
##     <td id="tableHTML_column_3">401</td>
##     <td id="tableHTML_column_4">-</td>
##     <td id="tableHTML_column_5">14.99</td>
##     <td id="tableHTML_column_6">401-B</td>
##     <td id="tableHTML_column_7">Format: Paperback, Language: Spanish</td>
##   </tr>
##   <tr>
##     <td id="tableHTML_rownames">16</td>
##     <td id="tableHTML_column_1"> Books</td>
##     <td id="tableHTML_column_2">Non-Fiction Guide</td>
##     <td id="tableHTML_column_3">402</td>
##     <td id="tableHTML_column_4">-</td>
##     <td id="tableHTML_column_5">24.99</td>
##     <td id="tableHTML_column_6">402-A</td>
##     <td id="tableHTML_column_7">Format: eBook, Language: English</td>
##   </tr>
##   <tr>
##     <td id="tableHTML_rownames">17</td>
##     <td id="tableHTML_column_1"> Books</td>
##     <td id="tableHTML_column_2">Non-Fiction Guide</td>
##     <td id="tableHTML_column_3">402</td>
##     <td id="tableHTML_column_4">-</td>
##     <td id="tableHTML_column_5">24.99</td>
##     <td id="tableHTML_column_6">402-B</td>
##     <td id="tableHTML_column_7">Format: Paperback, Language: French</td>
##   </tr>
##   <tr>
```

```
##   <td id="tableHTML_rownames">18</td>
##   <td id="tableHTML_column_1"> Sports Equipment</td>
##   <td id="tableHTML_column_2">Basketball</td>
##   <td id="tableHTML_column_3">501</td>
##   <td id="tableHTML_column_4">SportsGear</td>
##   <td id="tableHTML_column_5">29.99</td>
##   <td id="tableHTML_column_6">501-A</td>
##   <td id="tableHTML_column_7">Size: Size 7, Color: Orange</td>
## </tr>
## <tr>
##   <td id="tableHTML_rownames">19</td>
##   <td id="tableHTML_column_1"> Sports Equipment</td>
##   <td id="tableHTML_column_2">Tennis Racket</td>
##   <td id="tableHTML_column_3">502</td>
##   <td id="tableHTML_column_4">RacketPro</td>
##   <td id="tableHTML_column_5">89.99</td>
##   <td id="tableHTML_column_6">502-A</td>
##   <td id="tableHTML_column_7">Material: Graphite, Color: Black</td>
## </tr>
## <tr>
##   <td id="tableHTML_rownames">20</td>
##   <td id="tableHTML_column_1"> Sports Equipment</td>
##   <td id="tableHTML_column_2">Tennis Racket</td>
##   <td id="tableHTML_column_3">502</td>
##   <td id="tableHTML_column_4">RacketPro</td>
##   <td id="tableHTML_column_5">89.99</td>
##   <td id="tableHTML_column_6">502-B</td>
##   <td id="tableHTML_column_7">Material: Aluminum, Color: Silver</td>
## </tr>
## </tbody>
## </table>
```

```r
cat(lab7HTML)
```

```
##
## <table style="border-collapse:collapse;" class=table_4353 border=1>
## <thead>
## <tr>
##   <th id="tableHTML_header_1"> </th>
##   <th id="tableHTML_header_2">Category</th>
##   <th id="tableHTML_header_3">Item.Name</th>
##   <th id="tableHTML_header_4">Item.ID</th>
##   <th id="tableHTML_header_5">Brand</th>
##   <th id="tableHTML_header_6">Price</th>
##   <th id="tableHTML_header_7">Variation.ID</th>
##   <th id="tableHTML_header_8">Variation.Details</th>
## </tr>
## </thead>
## <tbody>
## <tr>
##   <td id="tableHTML_rownames">1</td>
##   <td id="tableHTML_column_1"> Electronics</td>
##   <td id="tableHTML_column_2">Smartphone</td>
##   <td id="tableHTML_column_3">101</td>
```

```
##     <td id="tableHTML_column_4">TechBrand</td>
##     <td id="tableHTML_column_5">699.99</td>
##     <td id="tableHTML_column_6">101-A</td>
##     <td id="tableHTML_column_7">Color: Black, Storage: 64GB</td>
## </tr>
## <tr>
##     <td id="tableHTML_rownames">2</td>
##     <td id="tableHTML_column_1"> Electronics</td>
##     <td id="tableHTML_column_2">Smartphone</td>
##     <td id="tableHTML_column_3">101</td>
##     <td id="tableHTML_column_4">TechBrand</td>
##     <td id="tableHTML_column_5">699.99</td>
##     <td id="tableHTML_column_6">101-B</td>
##     <td id="tableHTML_column_7">Color: White, Storage: 128GB</td>
## </tr>
## <tr>
##     <td id="tableHTML_rownames">3</td>
##     <td id="tableHTML_column_1"> Electronics</td>
##     <td id="tableHTML_column_2">Laptop</td>
##     <td id="tableHTML_column_3">102</td>
##     <td id="tableHTML_column_4">CompuBrand</td>
##     <td id="tableHTML_column_5">1099.99</td>
##     <td id="tableHTML_column_6">102-A</td>
##     <td id="tableHTML_column_7">Color: Silver, Storage: 256GB</td>
## </tr>
## <tr>
##     <td id="tableHTML_rownames">4</td>
##     <td id="tableHTML_column_1"> Electronics</td>
##     <td id="tableHTML_column_2">Laptop</td>
##     <td id="tableHTML_column_3">102</td>
##     <td id="tableHTML_column_4">CompuBrand</td>
##     <td id="tableHTML_column_5">1099.99</td>
##     <td id="tableHTML_column_6">102-B</td>
##     <td id="tableHTML_column_7">Color: Space Gray, Storage: 512GB</td>
## </tr>
## <tr>
##     <td id="tableHTML_rownames">5</td>
##     <td id="tableHTML_column_1"> Home Appliances</td>
##     <td id="tableHTML_column_2">Refrigerator</td>
##     <td id="tableHTML_column_3">201</td>
##     <td id="tableHTML_column_4">HomeCool</td>
##     <td id="tableHTML_column_5">899.99</td>
##     <td id="tableHTML_column_6">201-A</td>
##     <td id="tableHTML_column_7">Color: Stainless Steel, Capacity: 20 cu ft</td>
## </tr>
## <tr>
##     <td id="tableHTML_rownames">6</td>
##     <td id="tableHTML_column_1"> Home Appliances</td>
##     <td id="tableHTML_column_2">Refrigerator</td>
##     <td id="tableHTML_column_3">201</td>
##     <td id="tableHTML_column_4">HomeCool</td>
##     <td id="tableHTML_column_5">899.99</td>
##     <td id="tableHTML_column_6">201-B</td>
##     <td id="tableHTML_column_7">Color: White, Capacity: 18 cu ft</td>
```

```
## </tr>
## <tr>
##   <td id="tableHTML_rownames">7</td>
##   <td id="tableHTML_column_1"> Home Appliances</td>
##   <td id="tableHTML_column_2">Washing Machine</td>
##   <td id="tableHTML_column_3">202</td>
##   <td id="tableHTML_column_4">CleanTech</td>
##   <td id="tableHTML_column_5">499.99</td>
##   <td id="tableHTML_column_6">202-A</td>
##   <td id="tableHTML_column_7">Type: Front Load, Capacity: 4.5 cu ft</td>
## </tr>
## <tr>
##   <td id="tableHTML_rownames">8</td>
##   <td id="tableHTML_column_1"> Home Appliances</td>
##   <td id="tableHTML_column_2">Washing Machine</td>
##   <td id="tableHTML_column_3">202</td>
##   <td id="tableHTML_column_4">CleanTech</td>
##   <td id="tableHTML_column_5">499.99</td>
##   <td id="tableHTML_column_6">202-B</td>
##   <td id="tableHTML_column_7">Type: Top Load, Capacity: 5.0 cu ft</td>
## </tr>
## <tr>
##   <td id="tableHTML_rownames">9</td>
##   <td id="tableHTML_column_1"> Clothing</td>
##   <td id="tableHTML_column_2">T-Shirt</td>
##   <td id="tableHTML_column_3">301</td>
##   <td id="tableHTML_column_4">FashionCo</td>
##   <td id="tableHTML_column_5">19.99</td>
##   <td id="tableHTML_column_6">301-A</td>
##   <td id="tableHTML_column_7">Color: Blue, Size: S</td>
## </tr>
## <tr>
##   <td id="tableHTML_rownames">10</td>
##   <td id="tableHTML_column_1"> Clothing</td>
##   <td id="tableHTML_column_2">T-Shirt</td>
##   <td id="tableHTML_column_3">301</td>
##   <td id="tableHTML_column_4">FashionCo</td>
##   <td id="tableHTML_column_5">19.99</td>
##   <td id="tableHTML_column_6">301-B</td>
##   <td id="tableHTML_column_7">Color: Red, Size: M</td>
## </tr>
## <tr>
##   <td id="tableHTML_rownames">11</td>
##   <td id="tableHTML_column_1"> Clothing</td>
##   <td id="tableHTML_column_2">T-Shirt</td>
##   <td id="tableHTML_column_3">301</td>
##   <td id="tableHTML_column_4">FashionCo</td>
##   <td id="tableHTML_column_5">19.99</td>
##   <td id="tableHTML_column_6">301-C</td>
##   <td id="tableHTML_column_7">Color: Green, Size: L</td>
## </tr>
## <tr>
##   <td id="tableHTML_rownames">12</td>
##   <td id="tableHTML_column_1"> Clothing</td>
```

```
##     <td id="tableHTML_column_2">Jeans</td>
##     <td id="tableHTML_column_3">302</td>
##     <td id="tableHTML_column_4">DenimWorks</td>
##     <td id="tableHTML_column_5">49.99</td>
##     <td id="tableHTML_column_6">302-A</td>
##     <td id="tableHTML_column_7">Color: Dark Blue, Size: 32</td>
##   </tr>
##   <tr>
##     <td id="tableHTML_rownames">13</td>
##     <td id="tableHTML_column_1"> Clothing</td>
##     <td id="tableHTML_column_2">Jeans</td>
##     <td id="tableHTML_column_3">302</td>
##     <td id="tableHTML_column_4">DenimWorks</td>
##     <td id="tableHTML_column_5">49.99</td>
##     <td id="tableHTML_column_6">302-B</td>
##     <td id="tableHTML_column_7">Color: Light Blue, Size: 34</td>
##   </tr>
##   <tr>
##     <td id="tableHTML_rownames">14</td>
##     <td id="tableHTML_column_1"> Books</td>
##     <td id="tableHTML_column_2">Fiction Novel</td>
##     <td id="tableHTML_column_3">401</td>
##     <td id="tableHTML_column_4">-</td>
##     <td id="tableHTML_column_5">14.99</td>
##     <td id="tableHTML_column_6">401-A</td>
##     <td id="tableHTML_column_7">Format: Hardcover, Language: English</td>
##   </tr>
##   <tr>
##     <td id="tableHTML_rownames">15</td>
##     <td id="tableHTML_column_1"> Books</td>
##     <td id="tableHTML_column_2">Fiction Novel</td>
##     <td id="tableHTML_column_3">401</td>
##     <td id="tableHTML_column_4">-</td>
##     <td id="tableHTML_column_5">14.99</td>
##     <td id="tableHTML_column_6">401-B</td>
##     <td id="tableHTML_column_7">Format: Paperback, Language: Spanish</td>
##   </tr>
##   <tr>
##     <td id="tableHTML_rownames">16</td>
##     <td id="tableHTML_column_1"> Books</td>
##     <td id="tableHTML_column_2">Non-Fiction Guide</td>
##     <td id="tableHTML_column_3">402</td>
##     <td id="tableHTML_column_4">-</td>
##     <td id="tableHTML_column_5">24.99</td>
##     <td id="tableHTML_column_6">402-A</td>
##     <td id="tableHTML_column_7">Format: eBook, Language: English</td>
##   </tr>
##   <tr>
##     <td id="tableHTML_rownames">17</td>
##     <td id="tableHTML_column_1"> Books</td>
##     <td id="tableHTML_column_2">Non-Fiction Guide</td>
##     <td id="tableHTML_column_3">402</td>
##     <td id="tableHTML_column_4">-</td>
##     <td id="tableHTML_column_5">24.99</td>
```

```
##     <td id="tableHTML_column_6">402-B</td>
##     <td id="tableHTML_column_7">Format: Paperback, Language: French</td>
## </tr>
## <tr>
##     <td id="tableHTML_rownames">18</td>
##     <td id="tableHTML_column_1"> Sports Equipment</td>
##     <td id="tableHTML_column_2">Basketball</td>
##     <td id="tableHTML_column_3">501</td>
##     <td id="tableHTML_column_4">SportsGear</td>
##     <td id="tableHTML_column_5">29.99</td>
##     <td id="tableHTML_column_6">501-A</td>
##     <td id="tableHTML_column_7">Size: Size 7, Color: Orange</td>
## </tr>
## <tr>
##     <td id="tableHTML_rownames">19</td>
##     <td id="tableHTML_column_1"> Sports Equipment</td>
##     <td id="tableHTML_column_2">Tennis Racket</td>
##     <td id="tableHTML_column_3">502</td>
##     <td id="tableHTML_column_4">RacketPro</td>
##     <td id="tableHTML_column_5">89.99</td>
##     <td id="tableHTML_column_6">502-A</td>
##     <td id="tableHTML_column_7">Material: Graphite, Color: Black</td>
## </tr>
## <tr>
##     <td id="tableHTML_rownames">20</td>
##     <td id="tableHTML_column_1"> Sports Equipment</td>
##     <td id="tableHTML_column_2">Tennis Racket</td>
##     <td id="tableHTML_column_3">502</td>
##     <td id="tableHTML_column_4">RacketPro</td>
##     <td id="tableHTML_column_5">89.99</td>
##     <td id="tableHTML_column_6">502-B</td>
##     <td id="tableHTML_column_7">Material: Aluminum, Color: Silver</td>
## </tr>
## </tbody>
## </table>
```

Here we have saved the file locally, and also confirmed the table has been saved as HTML.

XML

I could not find any package that would simply convert our data into an XML file, so instead I wrote a function that would manually convert it into a more readable XML format.

```
xml_data <- xml_new_root("lab7data")
lab7data |>
  pmap(function(Category, Item.Name, Item.ID, Brand, Price, Variation.ID, Variation.Details) {
    node <- xml_add_child(xml_data, "Inventory")
    xml_add_child(node, "Category", Category)
    xml_add_child(node, "Item.Name", Item.Name)
    xml_add_child(node, "Item.ID", Item.ID)
    xml_add_child(node, "Brand", Brand)
    xml_add_child(node, "Price", Price)
    xml_add_child(node, "Variation.ID", Variation.ID)
    xml_add_child(node, "Variation.Details", Variation.Details)
  })
```

```
## [[1]]
## {xml_node}
## <Variation.Details>
##
## [[2]]
## {xml_node}
## <Variation.Details>
##
## [[3]]
## {xml_node}
## <Variation.Details>
##
## [[4]]
## {xml_node}
## <Variation.Details>
##
## [[5]]
## {xml_node}
## <Variation.Details>
##
## [[6]]
## {xml_node}
## <Variation.Details>
##
## [[7]]
## {xml_node}
## <Variation.Details>
##
## [[8]]
## {xml_node}
## <Variation.Details>
##
## [[9]]
## {xml_node}
## <Variation.Details>
##
## [[10]]
## {xml_node}
## <Variation.Details>
##
## [[11]]
## {xml_node}
## <Variation.Details>
##
## [[12]]
## {xml_node}
## <Variation.Details>
##
## [[13]]
## {xml_node}
## <Variation.Details>
##
## [[14]]
## {xml_node}
```

```
## <Variation.Details>
##
## [[15]]
## {xml_node}
## <Variation.Details>
##
## [[16]]
## {xml_node}
## <Variation.Details>
##
## [[17]]
## {xml_node}
## <Variation.Details>
##
## [[18]]
## {xml_node}
## <Variation.Details>
##
## [[19]]
## {xml_node}
## <Variation.Details>
##
## [[20]]
## {xml_node}
## <Variation.Details>
```

```r
write_xml(xml_data, "C:/Users/crumm/OneDrive/Documents/MSDS Spring 2025/DATA607 Labs/lab7data.xml")
read_xml_data <- read_xml("C:/Users/crumm/OneDrive/Documents/MSDS Spring 2025/DATA607 Labs/lab7data.xml")
cat(as.character(read_xml_data))
```

```
## <?xml version="1.0" encoding="UTF-8"?>
## <lab7data>
##   <Inventory>
##     <Category> Electronics</Category>
##     <Item.Name>Smartphone</Item.Name>
##     <Item.ID>101</Item.ID>
##     <Brand>TechBrand</Brand>
##     <Price>699.99</Price>
##     <Variation.ID>101-A</Variation.ID>
##     <Variation.Details>Color: Black, Storage: 64GB</Variation.Details>
##   </Inventory>
##   <Inventory>
##     <Category> Electronics</Category>
##     <Item.Name>Smartphone</Item.Name>
##     <Item.ID>101</Item.ID>
##     <Brand>TechBrand</Brand>
##     <Price>699.99</Price>
##     <Variation.ID>101-B</Variation.ID>
##     <Variation.Details>Color: White, Storage: 128GB</Variation.Details>
##   </Inventory>
##   <Inventory>
##     <Category> Electronics</Category>
##     <Item.Name>Laptop</Item.Name>
##     <Item.ID>102</Item.ID>
```

```
##      <Brand>CompuBrand</Brand>
##      <Price>1099.99</Price>
##      <Variation.ID>102-A</Variation.ID>
##      <Variation.Details>Color: Silver, Storage: 256GB</Variation.Details>
##    </Inventory>
##    <Inventory>
##      <Category> Electronics</Category>
##      <Item.Name>Laptop</Item.Name>
##      <Item.ID>102</Item.ID>
##      <Brand>CompuBrand</Brand>
##      <Price>1099.99</Price>
##      <Variation.ID>102-B</Variation.ID>
##      <Variation.Details>Color: Space Gray, Storage: 512GB</Variation.Details>
##    </Inventory>
##    <Inventory>
##      <Category> Home Appliances</Category>
##      <Item.Name>Refrigerator</Item.Name>
##      <Item.ID>201</Item.ID>
##      <Brand>HomeCool</Brand>
##      <Price>899.99</Price>
##      <Variation.ID>201-A</Variation.ID>
##      <Variation.Details>Color: Stainless Steel, Capacity: 20 cu ft</Variation.Details>
##    </Inventory>
##    <Inventory>
##      <Category> Home Appliances</Category>
##      <Item.Name>Refrigerator</Item.Name>
##      <Item.ID>201</Item.ID>
##      <Brand>HomeCool</Brand>
##      <Price>899.99</Price>
##      <Variation.ID>201-B</Variation.ID>
##      <Variation.Details>Color: White, Capacity: 18 cu ft</Variation.Details>
##    </Inventory>
##    <Inventory>
##      <Category> Home Appliances</Category>
##      <Item.Name>Washing Machine</Item.Name>
##      <Item.ID>202</Item.ID>
##      <Brand>CleanTech</Brand>
##      <Price>499.99</Price>
##      <Variation.ID>202-A</Variation.ID>
##      <Variation.Details>Type: Front Load, Capacity: 4.5 cu ft</Variation.Details>
##    </Inventory>
##    <Inventory>
##      <Category> Home Appliances</Category>
##      <Item.Name>Washing Machine</Item.Name>
##      <Item.ID>202</Item.ID>
##      <Brand>CleanTech</Brand>
##      <Price>499.99</Price>
##      <Variation.ID>202-B</Variation.ID>
##      <Variation.Details>Type: Top Load, Capacity: 5.0 cu ft</Variation.Details>
##    </Inventory>
##    <Inventory>
##      <Category> Clothing</Category>
##      <Item.Name>T-Shirt</Item.Name>
##      <Item.ID>301</Item.ID>
```

```
##      <Brand>FashionCo</Brand>
##      <Price>19.99</Price>
##      <Variation.ID>301-A</Variation.ID>
##      <Variation.Details>Color: Blue, Size: S</Variation.Details>
##    </Inventory>
##    <Inventory>
##      <Category> Clothing</Category>
##      <Item.Name>T-Shirt</Item.Name>
##      <Item.ID>301</Item.ID>
##      <Brand>FashionCo</Brand>
##      <Price>19.99</Price>
##      <Variation.ID>301-B</Variation.ID>
##      <Variation.Details>Color: Red, Size: M</Variation.Details>
##    </Inventory>
##    <Inventory>
##      <Category> Clothing</Category>
##      <Item.Name>T-Shirt</Item.Name>
##      <Item.ID>301</Item.ID>
##      <Brand>FashionCo</Brand>
##      <Price>19.99</Price>
##      <Variation.ID>301-C</Variation.ID>
##      <Variation.Details>Color: Green, Size: L</Variation.Details>
##    </Inventory>
##    <Inventory>
##      <Category> Clothing</Category>
##      <Item.Name>Jeans</Item.Name>
##      <Item.ID>302</Item.ID>
##      <Brand>DenimWorks</Brand>
##      <Price>49.99</Price>
##      <Variation.ID>302-A</Variation.ID>
##      <Variation.Details>Color: Dark Blue, Size: 32</Variation.Details>
##    </Inventory>
##    <Inventory>
##      <Category> Clothing</Category>
##      <Item.Name>Jeans</Item.Name>
##      <Item.ID>302</Item.ID>
##      <Brand>DenimWorks</Brand>
##      <Price>49.99</Price>
##      <Variation.ID>302-B</Variation.ID>
##      <Variation.Details>Color: Light Blue, Size: 34</Variation.Details>
##    </Inventory>
##    <Inventory>
##      <Category> Books</Category>
##      <Item.Name>Fiction Novel</Item.Name>
##      <Item.ID>401</Item.ID>
##      <Brand>-</Brand>
##      <Price>14.99</Price>
##      <Variation.ID>401-A</Variation.ID>
##      <Variation.Details>Format: Hardcover, Language: English</Variation.Details>
##    </Inventory>
##    <Inventory>
##      <Category> Books</Category>
##      <Item.Name>Fiction Novel</Item.Name>
##      <Item.ID>401</Item.ID>
```

```
##       <Brand>-</Brand>
##       <Price>14.99</Price>
##       <Variation.ID>401-B</Variation.ID>
##       <Variation.Details>Format: Paperback, Language: Spanish</Variation.Details>
##    </Inventory>
##    <Inventory>
##       <Category> Books</Category>
##       <Item.Name>Non-Fiction Guide</Item.Name>
##       <Item.ID>402</Item.ID>
##       <Brand>-</Brand>
##       <Price>24.99</Price>
##       <Variation.ID>402-A</Variation.ID>
##       <Variation.Details>Format: eBook, Language: English</Variation.Details>
##    </Inventory>
##    <Inventory>
##       <Category> Books</Category>
##       <Item.Name>Non-Fiction Guide</Item.Name>
##       <Item.ID>402</Item.ID>
##       <Brand>-</Brand>
##       <Price>24.99</Price>
##       <Variation.ID>402-B</Variation.ID>
##       <Variation.Details>Format: Paperback, Language: French</Variation.Details>
##    </Inventory>
##    <Inventory>
##       <Category> Sports Equipment</Category>
##       <Item.Name>Basketball</Item.Name>
##       <Item.ID>501</Item.ID>
##       <Brand>SportsGear</Brand>
##       <Price>29.99</Price>
##       <Variation.ID>501-A</Variation.ID>
##       <Variation.Details>Size: Size 7, Color: Orange</Variation.Details>
##    </Inventory>
##    <Inventory>
##       <Category> Sports Equipment</Category>
##       <Item.Name>Tennis Racket</Item.Name>
##       <Item.ID>502</Item.ID>
##       <Brand>RacketPro</Brand>
##       <Price>89.99</Price>
##       <Variation.ID>502-A</Variation.ID>
##       <Variation.Details>Material: Graphite, Color: Black</Variation.Details>
##    </Inventory>
##    <Inventory>
##       <Category> Sports Equipment</Category>
##       <Item.Name>Tennis Racket</Item.Name>
##       <Item.ID>502</Item.ID>
##       <Brand>RacketPro</Brand>
##       <Price>89.99</Price>
##       <Variation.ID>502-B</Variation.ID>
##       <Variation.Details>Material: Aluminum, Color: Silver</Variation.Details>
##    </Inventory>
## </lab7data>
```

Additionally, here we have saved the file, then called it. This makes sure our save was a success.

Pros and Cons

JSON - Some pros of are that data is stored in smaller size among small and medium datasets, and is often easy to read by actual people. Some cons are that it cannot support dates as data, as well as you cannot make comments within it.

Parquet - Storing data in columns rather than rows allows for more efficient data storage, therefore these files are typically smaller in size. Data can be queried faster as well due to being stored in columns. However, unlike JSON files, the data is not readable by humans.

HTML - This is the universal standard - all browsers support it and thus can be easily accessed. Data can also be presented in highly customizable tables. However, other languages such as JavaScript are needed to present information in more compelling ways. Additionally, quite a lot of coded is needed to store data that is simply stored in a CSV, or JSON file.

XML - Some pros are that it supports Unicode, and therefore can be accessed across languages. Additionally, you can perform changes to the data without affecting data presentation, and it is human readable. Unfortunately, it is fairly complex and this makes for larger sized files, and there everything is stored as text rather than any other data types.

##Conclusion