



By
Dr. Hussein Hazimeh

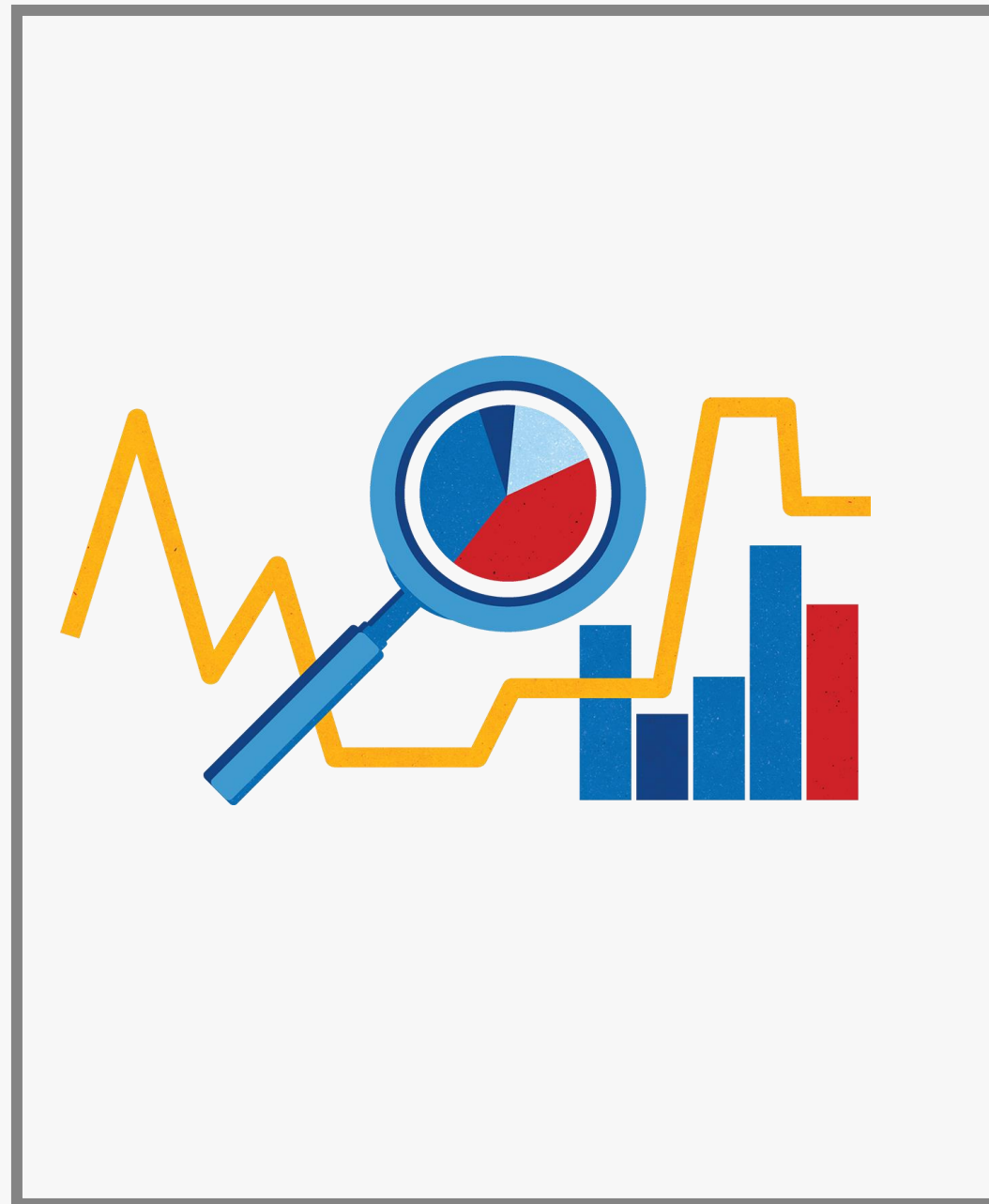
Lebanese University

Faculty of Information 1

Data Science Departement

2nd year – Data Analysis

March – 2020 – Chapter 2 – part 3



Agenda

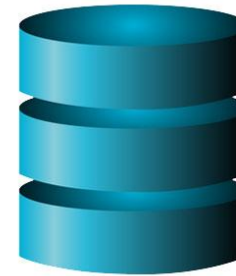
- » What is XML
- » What is JSON
- » XML vs JSON
- » XML & JSON limitations
- » Data scraping in XML
- » Data scraping in JSON
- » Assignment

Recap

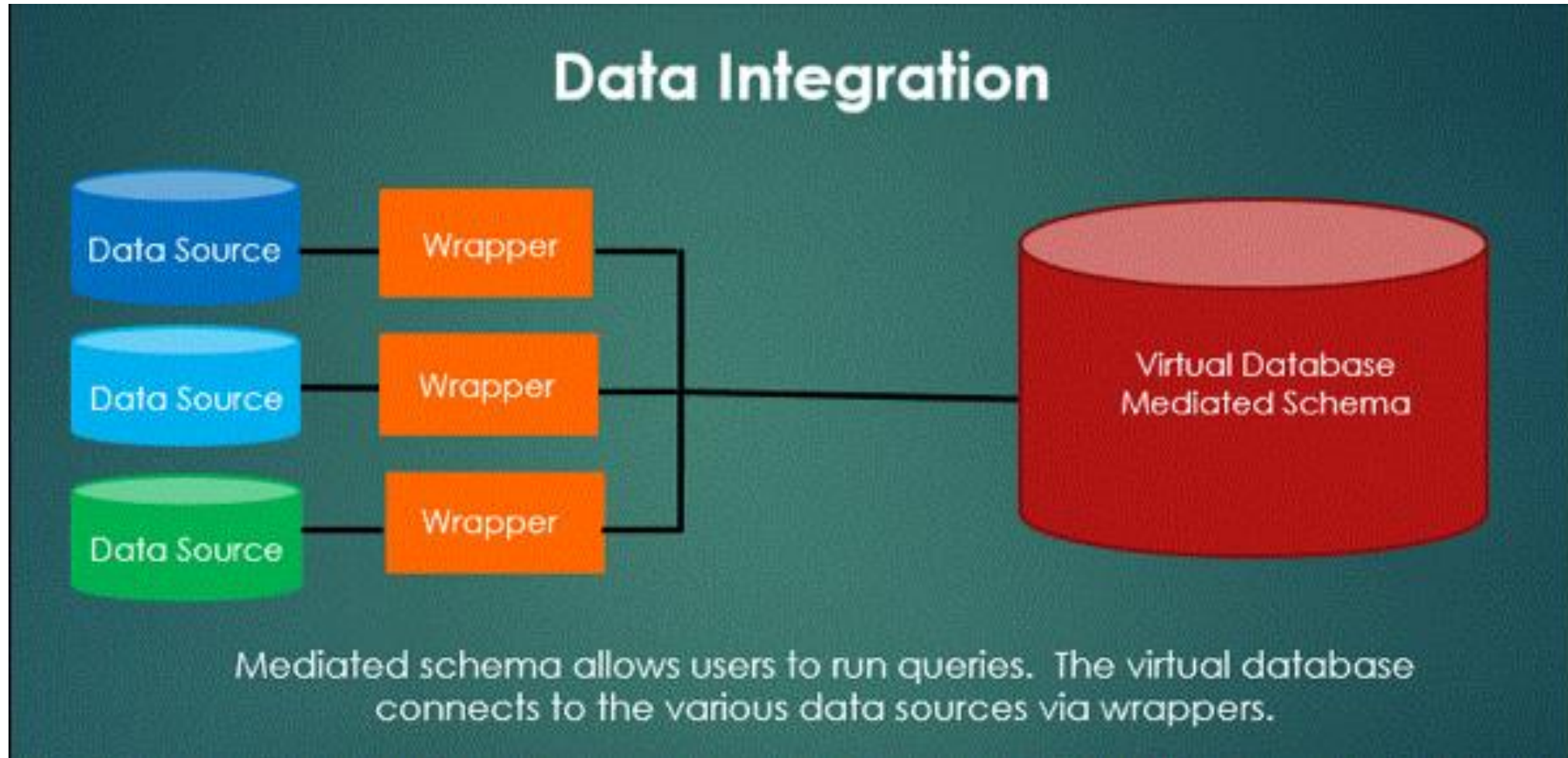
What is data integration?

- » Why data integration?
- » What are the different forms of data?
- » How can we extract different data using R?

Which data forms we learned?



Data integration schema





What is XML?

XML Defenition

» What is XML?

- » XML stands for eXtensible Markup Language
- » XML is a markup language much like HTML
- » XML was designed to store and transport data
- » XML was designed to be self-descriptive
- » XML is a W3C Recommendation
- » Developed in 1997

» XML Does Not DO Anything

- » Maybe it is a little hard to understand, but XML does not DO anything.

XML Example?

- » This message is a message to Hussein from Philippe, stored as XML:

```
<message>  
  <to>Philippe</to>  
  <from>Hussein</from>  
  <heading>Data Analysis</heading>  
  <body>Please find attached chapter 3.</body>  
</message>
```

- » The XML above is quite self-descriptive:

- It has sender information.
- It has receiver information
- It has a heading
- It has a message body.

- » But still, the XML above does not DO anything. XML is just information wrapped in tags.

XML vs HTML?

» What is the Difference Between XML and HTML?

» XML and HTML were designed with different goals

- XML was designed to carry data - with focus on what data is
- HTML was designed to display data - with focus on how data looks
- XML tags are not predefined like HTML tags are

» XML Does Not Use Predefined Tags

» The tags in the example above (like <to> and <from>) are not defined in any XML standard. These tags are "invented" by the author of the XML document.

» HTML works with predefined tags like <p>, <h1>, <table>, etc.

» With XML, the author must define both the tags and the document structure.

» xmltop[['from']]

JSON

» Whats is JSON?

- JSON: JavaScript Object Notation.
- JSON is a syntax for storing and exchanging data.
- JSON is text, written with JavaScript object notation.
- Developed in 2001

» Example

```
var myObj = {name: "John", age: 31, city: "New York"};  
var myJSON = JSON.stringify(myObj);  
window.location = "demo_json.php?x=" + myJSON;
```

» JSON Syntax Rules

» JSON syntax is derived from JavaScript object notation syntax:

- Data is in name/value pairs
- Data is separated by commas
- Curly braces hold objects
- Square brackets hold arrays

» JSON Data - A Name and a Value

- JSON data is written as name/value pairs.
- A name/value pair consists of a field name (in double quotes), followed by a colon, followed by a value:

```
{"name": "John"}
```

JSON vs XML

- » Both JSON and XML can be used to receive data from a web server.
- » The following JSON and XML examples both define an employees object, with an array of 3 employees:

JSON Example

```
{ "employees": [  
  { "firstName": "John", "lastName": "Doe" },  
  { "firstName": "Anna", "lastName": "Smith" },  
  { "firstName": "Peter", "lastName": "Jones" }  
]}
```

XML Example

```
<employees>  
  <employee>  
    <firstName>John</firstName> <lastName>Doe</lastName>  
  </employee>  
  <employee>  
    <firstName>Anna</firstName> <lastName>Smith</lastName>  
  </employee>  
  <employee>  
    <firstName>Peter</firstName> <lastName>Jones</lastName>  
  </employee>  
</employees>
```

JSON vs XML (cont.)

JSON	XML
Text based format (Not a Language)	Markup Language
Free to define anything	Has some rules
Smaller Size	Big in Size due to markups
JSON is similar to Java script Objects literals. Browser read faster.	Browser need parsers to handle XML. Slow processing.
No support on namespaces and comments	Both are supported.

XML & JSON Limitations

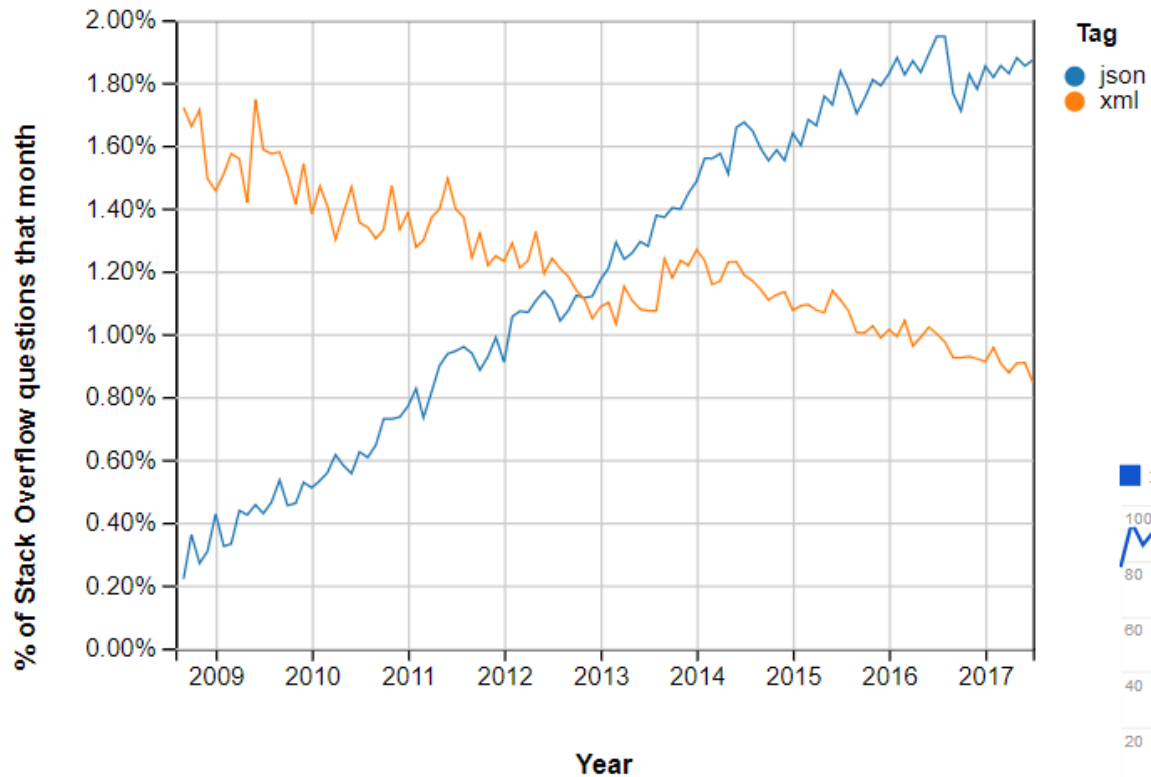
» **What are the limitation of XML?**

- » XML is not optimized for access speed.
- » XML document is less readable compared to other text-based data transmission formats such as JSON.
- » XML doesn't support array.
- » XML file sizes are usually very large due to its verbose nature, it is totally dependant on who is writing it

» **What are the limitation of JSON?**

- » JSON has no error handling for JSON calls. If the dynamic script insertion works, you get called and will get the response perfectly.
- » If not inserted, nothing happens. It just fails silently.
- » Another major drawback of JSON is that it can be quite dangerous if used with untrusted services or untrusted browsers, because a JSON service returns a JSON response wrapped in a function call, which will be executed by the browser if it will be used with untrusted browser it can be hacked, this makes the hosting Web Application Vulnerable to a variety of attacks.

XML & JSON graphs



XML vs JSON

