

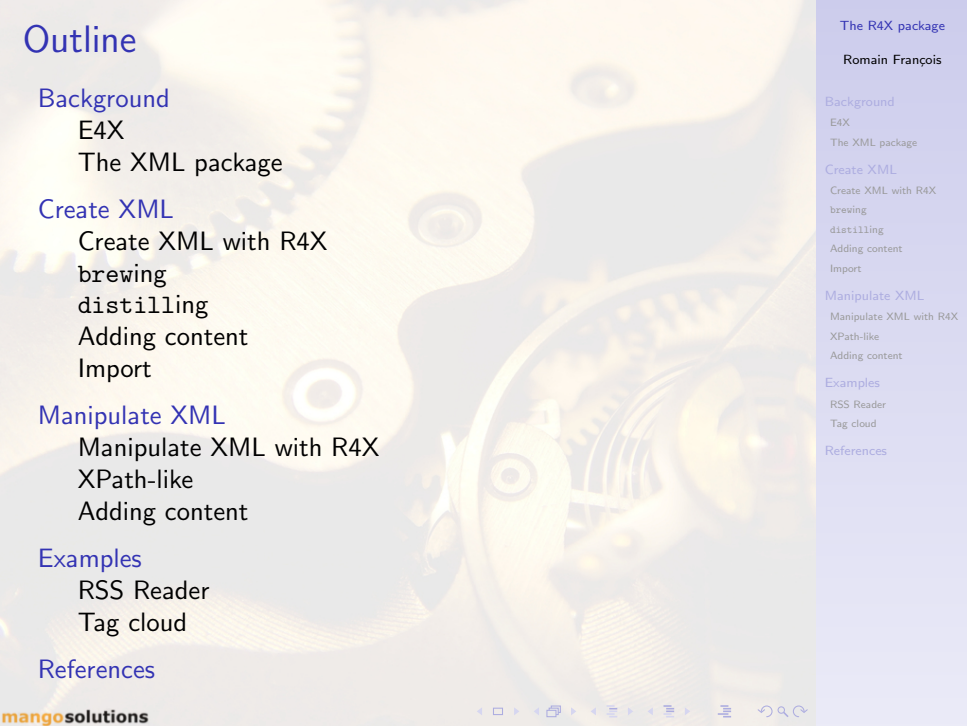
A close-up, macro photograph of the intricate mechanical gears and components of a watch movement. The metal parts are polished and show fine textures. The lighting is warm and directional, creating strong highlights and deep shadows that emphasize the three-dimensional nature of the gears.

The R4X package

Convenient XML Manipulation for R

Romain François

mangosolutions
data analysis that delivers




Outline

- Background
 - E4X
 - The XML package
- Create XML
 - Create XML with R4X
 - brewing
 - distilling
 - Adding content
 - Import
- Manipulate XML
 - Manipulate XML with R4X
 - XPath-like
 - Adding content
- Examples
 - RSS Reader
 - Tag cloud
- References

The R4X package

Romain François

- Background
 - E4X
 - The XML package
- Create XML
 - Create XML with R4X
 - brewing
 - distilling
 - Adding content
 - Import
- Manipulate XML
 - Manipulate XML with R4X
 - XPath-like
 - Adding content
- Examples
 - RSS Reader
 - Tag cloud
- References



mango**solutions**

Romain Francois

E4X

The XML package

brewing

Manipulate XML with R4X

Adding content

RSS Reader

Tag cloud

<http://en.wikipedia.org/wiki/E4X>

```

alert( sales.item.(@type == "carrot").@quantity );
alert( sales.@vendor );
for each( var price in sales..@price )
    alert( price );

```

The XML Package

from the $\hat{\Omega}$ project. <http://www.omegahat.org/RXML/>

This package provides facilities for the S language to

- ▶ parse XML files, URLs and strings, using either the DOM (Document Object Model)/tree-based approach, or the event-driven SAX (Simple API for XML) mechanism;
- ▶ parse HTML documents,
- ▶ perform XPath queries on a document,
- ▶ generate XML content to buffers, files, URLs, and internal XML trees;
- ▶ read DTDs as S objects.

slicing with [

The *single* square bracket [gives an XMLNode or a list of XMLNode if the path matches more than one node

```
> x[ "child" ] # mutiple <child> : list of XMLNode
```

\$child

<child id="1">

```
<subchild id="sub1">foo</subchild>
```

```
<subchild id="sub2">bar</subchild>
```

</child>

```
> x[ "fruits" ] # single <fruits> : XMLNode
```

```
$fruits
```

```
<fruits>
```

```
<fruit>banana</fruit>
```

```
<fruit>mango</fruit>
```

</fruits>

RSS: R Site Summary

Definition of RSS from the w3c. See <http://www.w3schools.com/rss> for more information.

- ▶ RSS stands for Really Simple Syndication
- ▶ RSS allows you to syndicate your site content
- ▶ RSS defines an easy way to share and view headlines and content
- ▶ RSS files can be automatically updated
- ▶ RSS allows personalized views for different sites
- ▶ RSS is written in XML

Example RSS feed from <http://www.w3schools.com/rss>.

Example RSS feed from <http://www.w3schools.com/rss>.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<rss version="2.0">
<channel>
  <title>W3Schools Home Page</title>
  <link>http://www.w3schools.com</link>
  <description>
    Free web building tutorials
  </description>
  <item>
    <title>RSS Tutorial</title>
    <link>http://www.w3schools.com/rss</link>
    <description>
      New RSS tutorial on W3Schools
    </description>
  </item>
</channel>
</rss>
```


1 2 al algorithm allows analyses **analysis** applications applied approach arbitrary association available basic bayesian
binary book bootstrap c calculate calculation carlo censored chain class classes classification cluster clustering code collection
common components computation computational compute computing conditional confidence control correlation count covariates
create currently curves **data** database datasets density described design designed detection different discrete display distance
distribution eg either engineering environment error estimate estimating estimation estimator et etc exact examples
experiments features file finance financial first fit fitting framework **function** functionality gaussian gene general
generalized genetic graph graphical graphics group gui hazard hierarchical if implementation implemented implements include
included including independent inference information interface intervals its kernel large level library likelihood linear local
logistic main manipulating map markov matrices matrix maximum may mean measures **method** microarray missing mixture
model modeling modelling monte most multiple multivariate network nonlinear nonparametric normal number object
observations order output **package** parameter parametric perform plot plotting point population possible power
probability problems procedure process processes program programming proportional **provide** provided quantitative **r** random
regression related response results risk robust routines s sample sampling selection series set simple simulation single
smoothing so software spatial specified splus squares standard statistical statistics structure support survival system teaching test
testing theory through time tools trees univariate useful user uses **using** utilities utility value **variable** variance various vector
version very wavelet weighted without written



Questions ?

rfrancois@mango-solutions.com

mangosolutions
data analysis that delivers