

# Assignment 2

*Aaron Niskin*

*August 25, 2016*

There is a great deal of data available in the form of Extensible Markup Language (xml) files. There is a good reference for xml files: <https://www.sitepoint.com/really-good-introduction-xml/> Navigate to <http://www.cs.washington.edu/research/xmldatasets/www/repository.html#auctions> Prepare an R Markdown document which documents the use of R tools/code to address the following

1. Download the 277KB xml file associated to Reed courses.

```
library(XML)
reed_df <- xmlToDataFrame("http://www.cs.washington.edu/research/xmldatasets/data/courses/reed.xml")
```

2. How many distinct subjects are listed in the document?

```
answer_2 <- length(unique(reed_df[, 'subj']))
answer_2
```

```
## [1] 31
```

3. How many courses have a NULL instructor listing?

```
answer_3 <- length(reed_df[reed_df$instructor == '', ], 'instructor')
answer_3
```

```
## [1] 15
```

4. How many distinct instructors are listed?

```
answer_4 <- length(unique(reed_df[, 'instructor'] [reed_df$instructor != '']))
answer_4
```

```
## [1] 135
```

There are 135 instructors.