Data scraping

Nicholas Horton (nhorton@amherst.edu)

October 7, 2014

```
require(XML)
require(mosaic)
require(lubridate)
wikipedia = "http://en.wikipedia.org/wiki"
page = "List_of_nuclear_reactors"
result = readHTMLTable(paste(wikipedia, page, sep="/"),
  stringsAsFactors=FALSE)
table = result[[21]]  # change to appropriate table number
names(table)
```

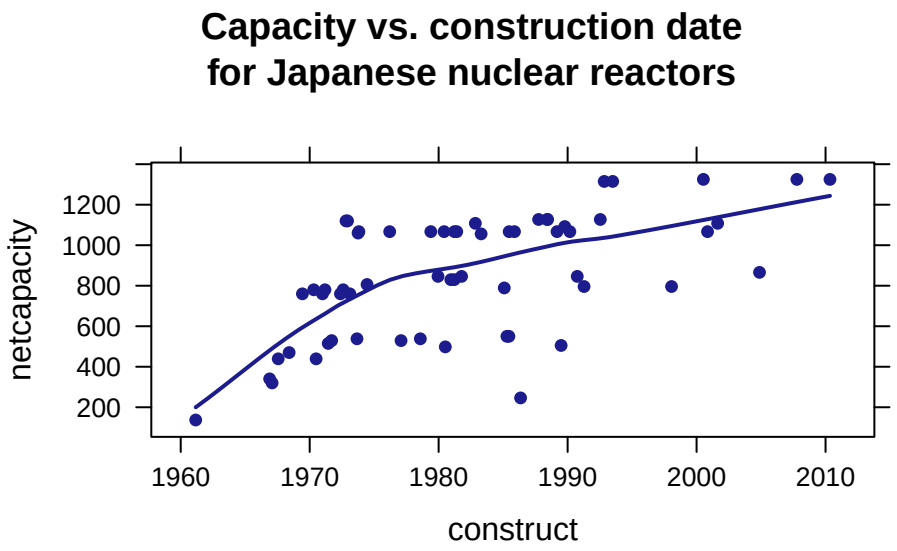
```
##  [1] "V1"  "V2"  "V3"  "V4"  "V5"  "V6"  "V7"  "V8"  "V9"  "V10"
```

```
finaltable = mutate(table,
      netcapacity = as.numeric(V6),
      status = V5,
      construct = dmy(V8))  # from lubridate
```

```
tally(~ status, data=finaltable)
```

```
##
##              Planned             Shutdown Suspended Operation
##                    3                   10                   48
##  Under Construction                 <NA>
##                    2                    1
```

```
xyplot(netcapacity ~ construct, type=c("p", "smooth"),
       main = "Capacity vs. construction date\nfor Japanese nuclear reactors",
       data=finaltable)
```



Interpretation: the average net capacity of nuclear power plants in Japan tended to increase over time (but then plateaued in recent years).

YOUR MISSION

In groups of 3, find an interesting Wikipedia page with a table, scrape the data from it, and generate a figure that tells an interesting story. (I'd suggest finding a simple table, as more complex ones are harder to parse.)

You should include a sentence which interprets the figure.