Stat 133, Fall 2014
Homework 7: KML
Due Thursday, November 13 at 11:55pm on bSpace

Your task is to create a Google Earth display of the infant mortality and population for countries around the world. Google Earth uses a particular dialect of XML called KML to describe geographic data. You are to merge two data sets together and then create a KML file using the resulting data frame. You will turn in an R script file with the code that creates the KML file. The KML file has a basic structure that looks something like this:

```
<?xml version="1.0"?>

<kml xmlns="http://www.opengis.net/kml/2.2">

  <Document>

    <name>Country Facts</name>

    <description>Infant mortality</description>
    <LookAt>
      <longitude>-121</longitude>
      <latitude>43</latitude>
      <altitude>4100000</altitude>
      <tilt>0</tilt>
      <heading>0</heading>
      <altitudeMode>absolute</altitudeMode>
    </LookAt>
<Folder>
     <name>CIA Fact Book</name>
    <Placemark id="US">
       <name>United States</name>
     <Point>
        <coordinates>-147.426, 60.929, 0</coordinates>
     </Point>
   </Placemark>

 ...

    </Folder>
  </Document>
</kml>
```

A few notes:

- To find out about valid tags in KML, you can refer to the KML reference at `http://code.google.com/apis/kml/documentation/kmlreference.html`.
- Don't try to do everything at once; think about what the XML code is that you want to create first, and only then write your R code to generate it. Draw a tree diagram to help you translate between one and the other.
- When creating XML/KML objects in `R`, be careful about the order of the commands, as this can affect the finished document.

PART 1. Create the data frame. Create a data frame that contains the various pieces of information, merging data from two sources:

The 2011 CIA Fact Book, which is available in an XML format at
`http://www.stat.berkeley.edu/users/nolan/stat133/data/factbook.xml`
This file contains the country populations and infant mortality rates. Be careful though, the information about population appears in a different order than the information about mortality rates (and for different countries). If you are having trouble locating the infant mortality data, try opening the XML file in a plain text editor such as NotePad or TextEdit. It's a very large file. Search the file for Infant until you find the node that has infant mortality rates. You can also do that for population although the population table is about the 23rd occurrence of the word population in this file. Examine the node and its children carefully to figure out how to extract the information that you want. The table in the appendix of the fact book contains county codes and country name. There's a line in the .R file that will find it for you.

The second dataset is for latitude and longitude of the country centers, given in HW7.rda. This is originally from http://www.maxmind.com but I've processed it so that it has country codes that match with the country codes you'll get from the CIA data frame. Merge this with the data frame from the CIA as described in HW7.R.

PART 2. Create the KML document.

1. Write a script to create the first few levels of the KML hierarchy, e.g., the root node, the Document node and its name, description, and LookAt nodes.
2. Augment your script to add Placemarks using the function in HW7.R.

   Use a for loop to add a Placemark node for each country in the data frame. Notice that the country code is the id attribute on Placemark, the country name is the name element, and the description includes the facts about the country. Save this file with *saveXML* and open it in Google Earth. You should see a yellow pushpin for each country, and when you click on the pushpin, the stats for the country appear.