

Nov 19, 09 12:05

make_fake_country_groups.py

Page 1/1

```
#!/ python

""" This program splits CIA World Factbook country data into
two groups. The split is random, so there ought not to be any
statistical significant differences between the groups.
"""

#Get a library of random number generators:
import random

# Open the input file:
input = open("data6/country_areas2.txt", "r")

for (i,line) in enumerate(input):

    # Python reads in the "newline" character at the end of each line.
    # It's most convenient to throw it away before we do any processing:
    line = line.rstrip()

    if line.startswith('#'):
        # Drop comments
        pass
    elif i == 0:
        # Replace the first line with the variable names:
        print "i,Name,area,group"
    elif line == '':
        # The last line is empty. We'll just eat it.
        pass
    else:
        # Split the line into three fields at a tab:
        n, name, area = line.split('\t')

        # Add the fake group information:
        group = random.random() < 0.5

        # We need to do this because the "area" field is
        # filled with spaces. It looks pretty strange if we print
        # it directly into a comma-separated-value format.
        area = int(area)

        # Print it out:
        print "%s,%s,%d,%s" % (n, name, area, group)
```