

Introduction to Spatial Statistics

Shonda Kuiper

Thursday, July 10, 2014

In this lab, you will be asked to find patterns and draw conclusions using spatial data collected from the CIA World Fact Book, <https://www.cia.gov/library/publications/the-world-factbook/rankorder/rankorderguide.html>.

```
require(DCFdevel)
require(dplyr)
require(mosaic)
# You can get DCFdevel by: `install_github( dtkaplan/DCFdevel )`
```

Our dataset consists of four variables:

- **country**: the name of each country within our dataset
- **totalmilitary**: Total annual military expenditures
- **GDP**: (purchasing power parity) compares the gross domestic product (GDP) or value of all final goods
- **pop**: estimated population of that country

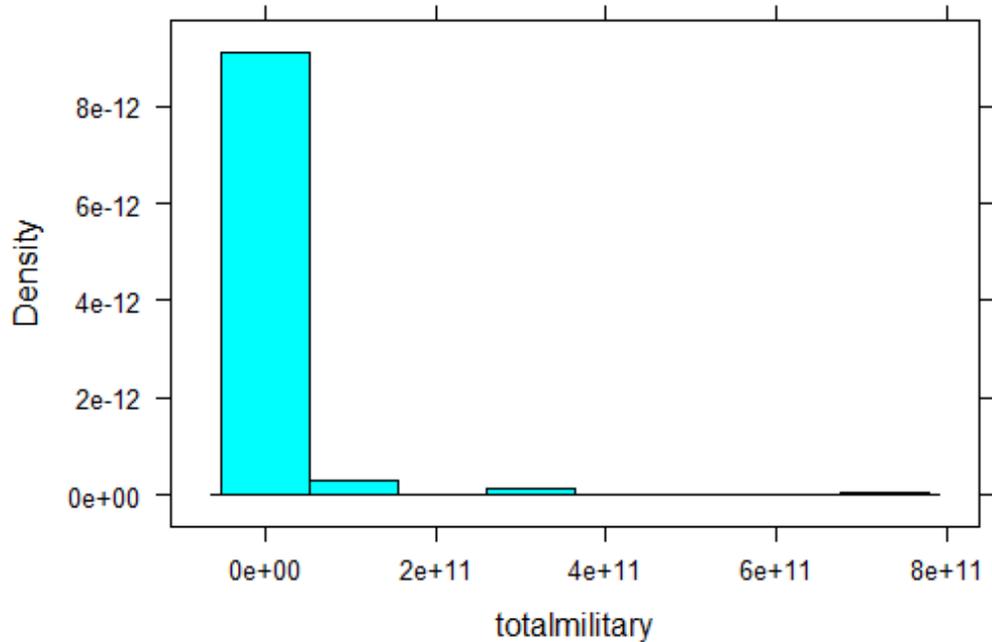
The first five rows of data are shown below:

```
head(CIAMilitary,5)
##      country      GDP totalmilitary      pop
## 1  Albania 2.834e+10  4.166e+08  3020209
## 2  Algeria 2.847e+11  1.275e+10 38813722
## 3   Angola 1.318e+11  4.784e+09 19088106
## 4 Argentina 7.710e+11  7.016e+09 43024374
## 5  Armenia 2.061e+10  8.079e+08  3060631
```

1. The following code creates a histogram of the total annual military expenditures for each country. What useful information does this graph provide?

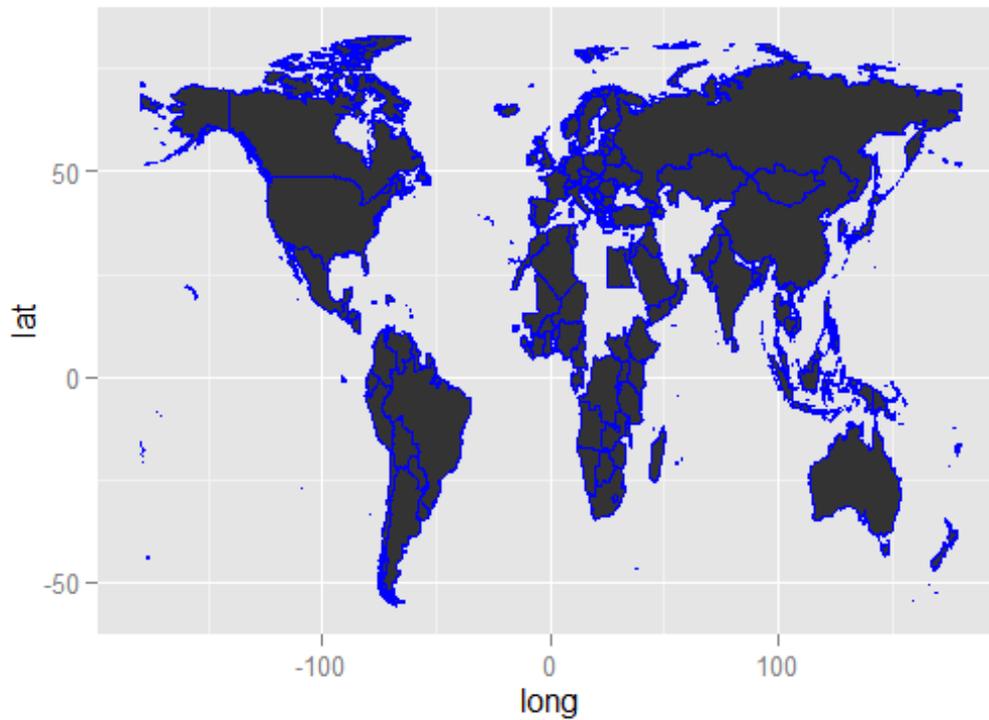
```
histogram(~totalmilitary, data = CIAMilitary, main = "Total Annual Military Spending for all Countries")
```

Total Annual Military Spending for all Countries

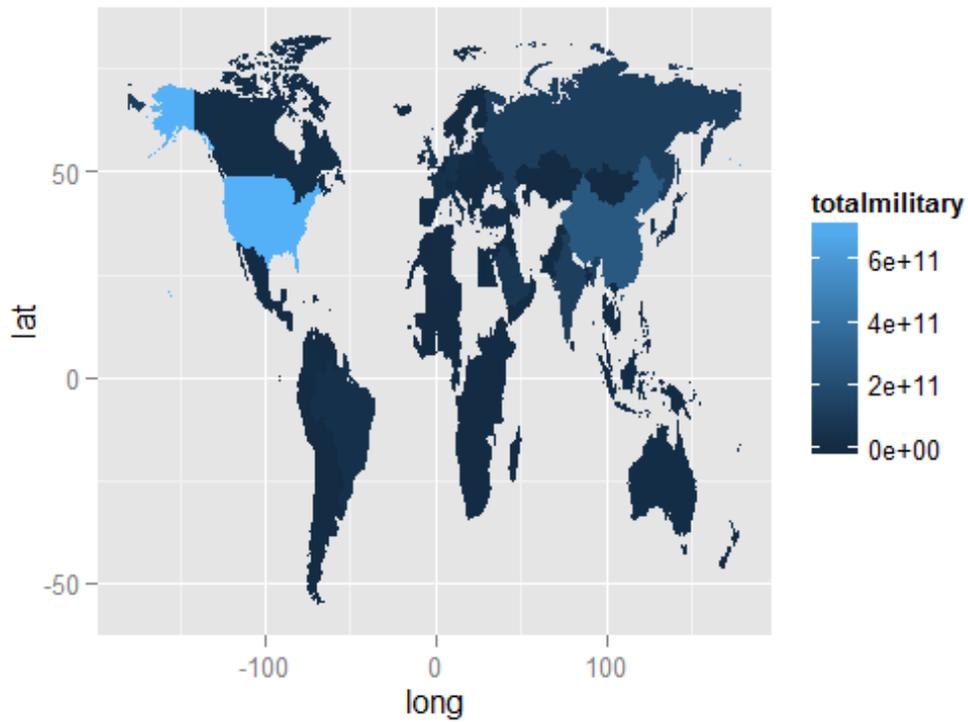


2. The following code creates a spatial graph to view the same information. Notice that the first two lines of code make a base map and the last line of the code chooses which term will be plotted. What useful information does this graph provide?

```
map = makeWorldMap(CIAMilitary, "country")  
map + geom_polygon(color = "blue")
```



```
map + geom_polygon(aes(fill=totalmilitary))
```



3. Describe how both graphs demonstrate that the data is highly skewed.

4. The following code allows you to sort the CIAmilitary data by totalmilitary. It then provides the top three and last three rows of data. Which country spends the most on annual military expenditures?

```
CIAmilitary = arrange(CIAmilitary, totalmilitary)
head(CIAmilitary, 3)
```

```
##      country      GDP totalmilitary    pop
## 1    Kuwait 1.658e+11           0 2742711
## 2    Iceland 1.311e+10      17043000 317351
## 3 Seychelles 2.404e+09      21395600  91650
```

```
tail(CIAmilitary, 3)
```

```
##      country      GDP totalmilitary    pop
## 130 European Union 1.585e+13    2.615e+11 5.114e+08
## 131      China 1.339e+13    2.665e+11 1.356e+09
## 132 United States 1.672e+13    7.273e+11 3.189e+08
```

5. Notice that we have a zero value, either Kuwait has no military spending or we simply do not have data for this country. Conduct a quick on-line search to determine which of these are true. For this lab, we will simply remove this one point.

```
CIAmilitary = filter(CIAmilitary, country != "Kuwait")
head(CIAmilitary, 3)
```

```
##      country      GDP totalmilitary    pop
## 1    Iceland 1.311e+10      17043000 317351
## 2 Seychelles 2.404e+09      21395600  91650
## 3    Liberia 2.898e+09      23763600 4092310
```

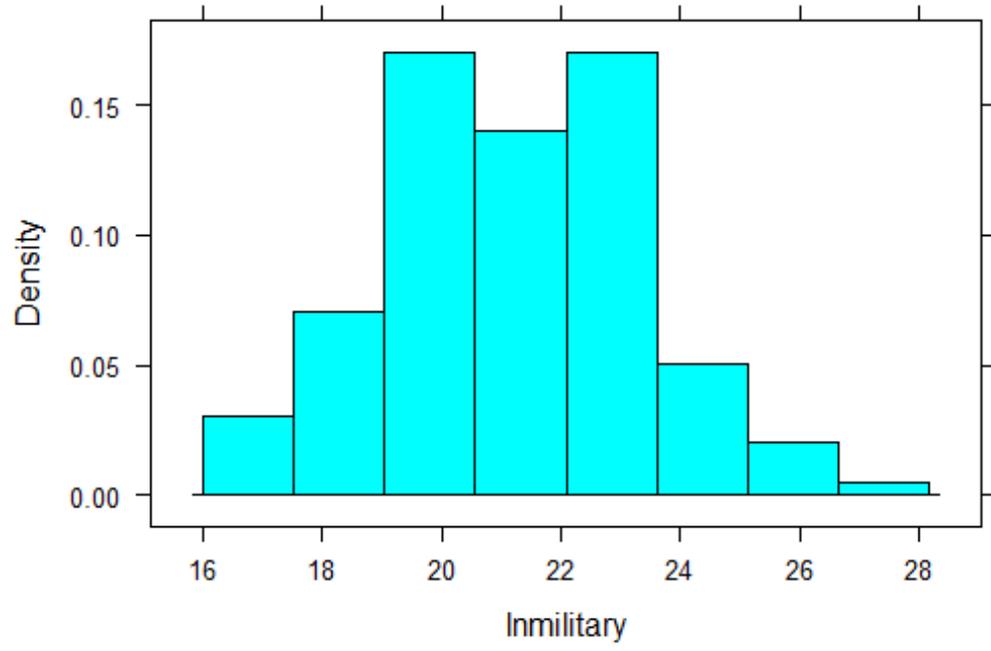
```
tail(CIAmilitary, 3)
```

```
##      country      GDP totalmilitary    pop
## 129 European Union 1.585e+13    2.615e+11 5.114e+08
## 130      China 1.339e+13    2.665e+11 1.356e+09
## 131 United States 1.672e+13    7.273e+11 3.189e+08
```

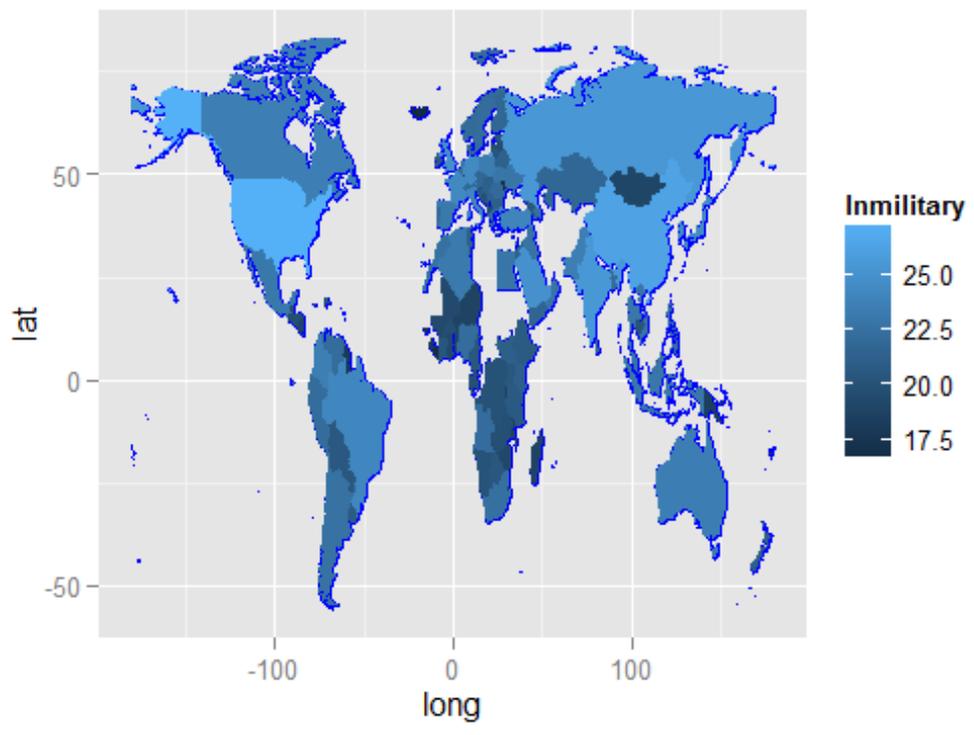
6. The following code creates a new variable, the natural log of military spending. The lnmilitary variable is then used to create two new graphs. Which variable, military or lnmilitary is more useful?

```
CIAmilitary = mutate(CIAmilitary, lnmilitary = log(totalmilitary+1))
histogram(~lnmilitary, data = CIAmilitary, main = "Log of Annual Military Spending for all Countries")
```

Log of Annual Military Spending for all Countries



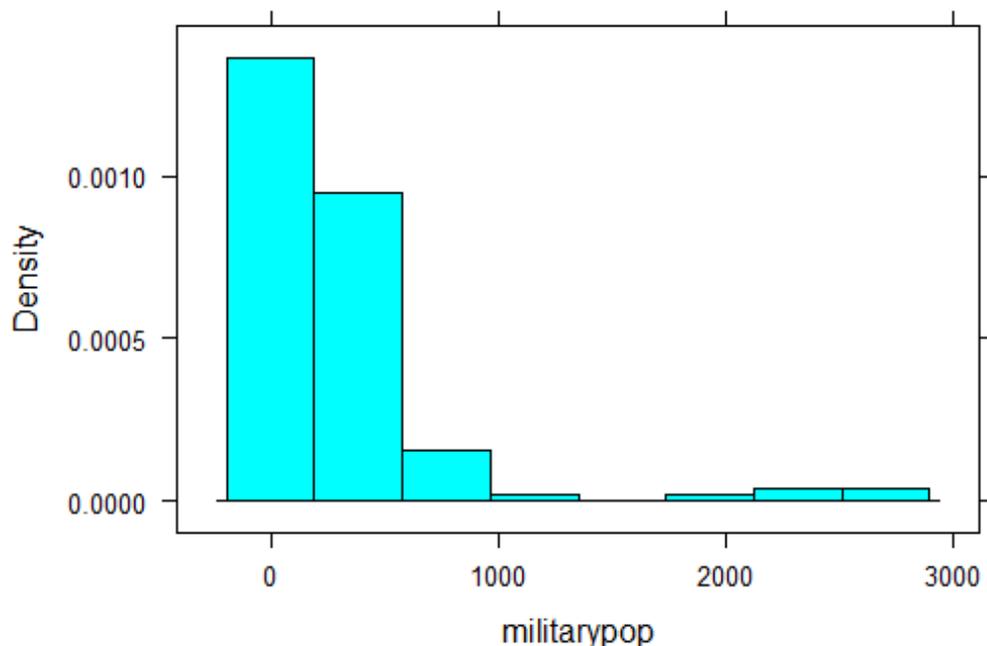
```
map = makeWorldMap(CIAMilitary, "country")
map = map + geom_polygon(color = "blue")
map + geom_polygon(aes(fill=lnmilitary))
```



7. Clearly larger countries with more people are likely to spend more money on military expenditures. Instead of using `totalmilitary`, the following code plots military spending per capita. What do these graphs tell you? which country spends the most on military per capita?

```
CIAmilitary = mutate(CIAmilitary, militarypop = totalmilitary/(pop))  
histogram(~militarypop, data = CIAmilitary, main = "Annual Military Spending  
per Capita")
```

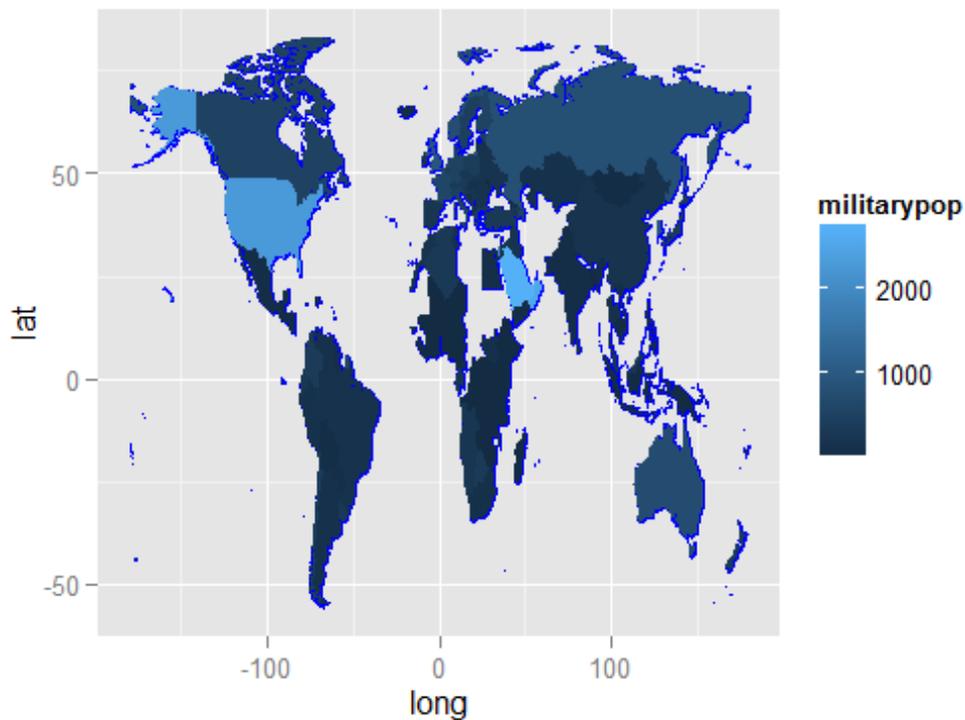
Annual Military Spending per Capita



```
CIAmilitary = arrange(CIAmilitary, militarypop)  
tail(CIAmilitary, 4)
```

##	country	GDP	totalmilitary	pop	lnmilitary	militarypop
## 128	Singapore	3.390e+11	1.193e+10	5567301	23.20	2143
## 129	United States	1.672e+13	7.273e+11	318892103	27.31	2281
## 130	Oman	9.486e+10	8.167e+09	3219775	22.82	2537
## 131	Saudi Arabia	9.278e+11	7.404e+10	27345986	25.03	2707

```
map = makeWorldMap(CIAmilitary, "country")  
map = map + geom_polygon(color = "blue")  
map + geom_polygon(aes(fill=militarypop))
```



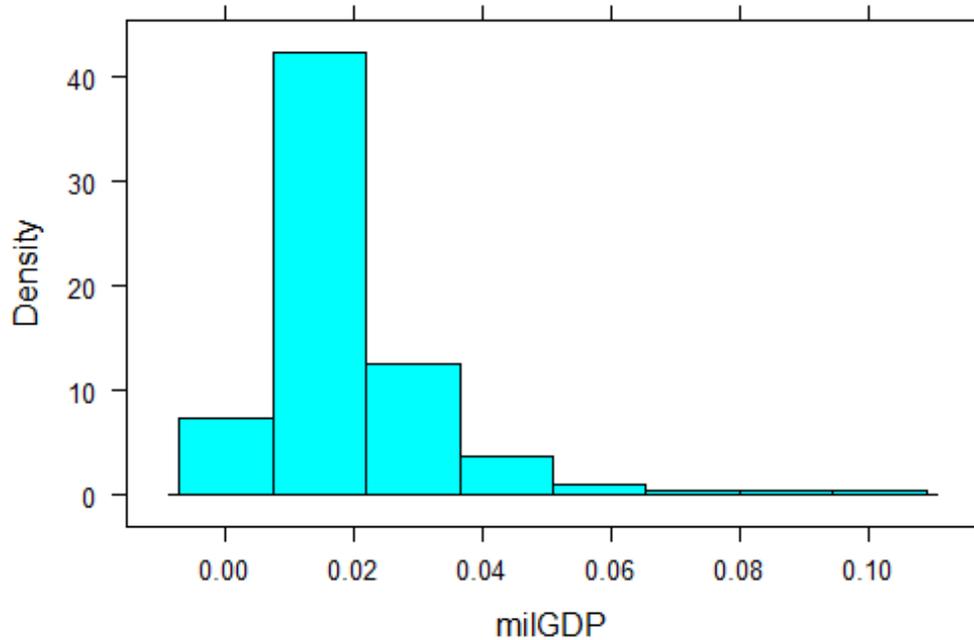
8. We can also look at military spending as a percentage of the countries economy (we will use GDP). Modify the above code to create a variable called milGDP. Create graphs and interpret the results. Are there any surprising results?

```

CIAmilitary = mutate(CIAmilitary, milGDP = totalmilitary/(GDP))
histogram(~milGDP, data = CIAmilitary, main = "Annual Military Spending by GDP")

```

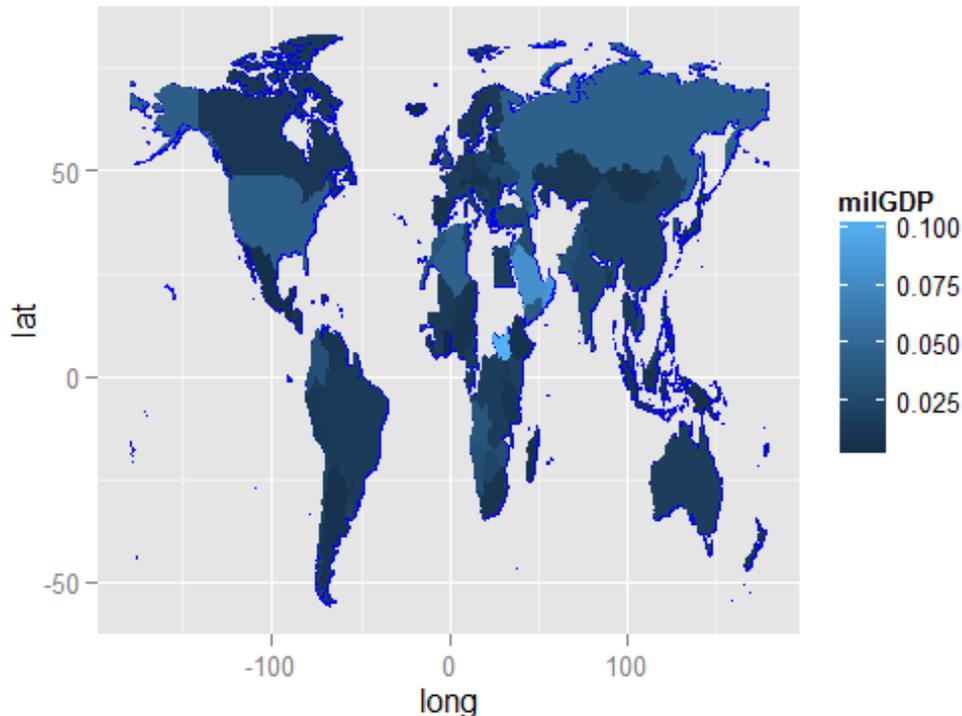
Annual Military Spending by GDP



```
CIAmilitary = arrange(CIAmilitary, milGDP)
tail(CIAmilitary, 10)
```

##	country	GDP	totalmilitary	pop	lnmilitary	militarypop
## 122	Lebanon	6.431e+10	2.598e+09	5882562	21.68	441.7
## 123	United States	1.672e+13	7.273e+11	318892103	27.31	2280.8
## 124	Russia	2.553e+12	1.141e+11	142470272	25.46	801.0
## 125	Algeria	2.847e+11	1.275e+10	38813722	23.27	328.6
## 126	Jordan	4.002e+10	1.861e+09	7930491	21.34	234.7
## 127	Azerbaijan	1.027e+11	5.340e+09	9686210	22.40	551.3
## 128	Israel	2.732e+11	1.555e+10	7821850	23.47	1987.4
## 129	Saudi Arabia	9.278e+11	7.404e+10	27345986	25.03	2707.5
## 130	Oman	9.486e+10	8.167e+09	3219775	22.82	2536.7
## 131	South Sudan	1.471e+10	1.518e+09	11562695	21.14	131.3
##	milGDP					
## 122	0.0404					
## 123	0.0435					
## 124	0.0447					
## 125	0.0448					
## 126	0.0465					
## 127	0.0520					
## 128	0.0569					
## 129	0.0798					
## 130	0.0861					
## 131	0.1032					

```
map = makeWorldMap(CIAMilitary, "country")
map = map + geom_polygon(color = "blue")
map + geom_polygon(aes(fill=milGDP))
```



9. All four military spending variables (`military`, `lnmilitary`, `militarypop`, and `militaryGDP`) used in this lab are accurate, but they provide very different results. If you are a liberal who wants to reduce U. S. Military spending, which graph do you think would be most effective in supporting your views? If you were a conservative wanting to increase U. S. Military spending which graph do you think would most effectively support your views? Can you determine which variable would be best to truly understand military spending in each country?

Your Take Home Assignment(Day 1)

The `CIAMilitary2014` dataset contains many more country variables. Your group assignment is to find three to four variables that you believe will create interesting graphs. Use the code provided above to create graphs. You will have time during our next class period to get help with any programming questions and assisting with interpreting your results. Before class answer the following questions.

It would be nice to use the `?CIAMilitary2014` file for the data?

10. What combination of variables are you interested in evaluating?
11. After looking at the data (hopefully creating a few plots), do you expect any transformations should be made on your variables?

12. *Extra Credit Question:* Find an example of a graphic on the web or in print that may have selected variables that may be biased towards a particular idea.

Your Take Home Assignment (Day 2)

Modify this current R Markdown file to create a one page pdf document that: * States the question you are researching. * Shows any modifications you made on the dataset. * Includes one or two graphics. * Uses the graphs to justify your conclusions.