Problem set 4

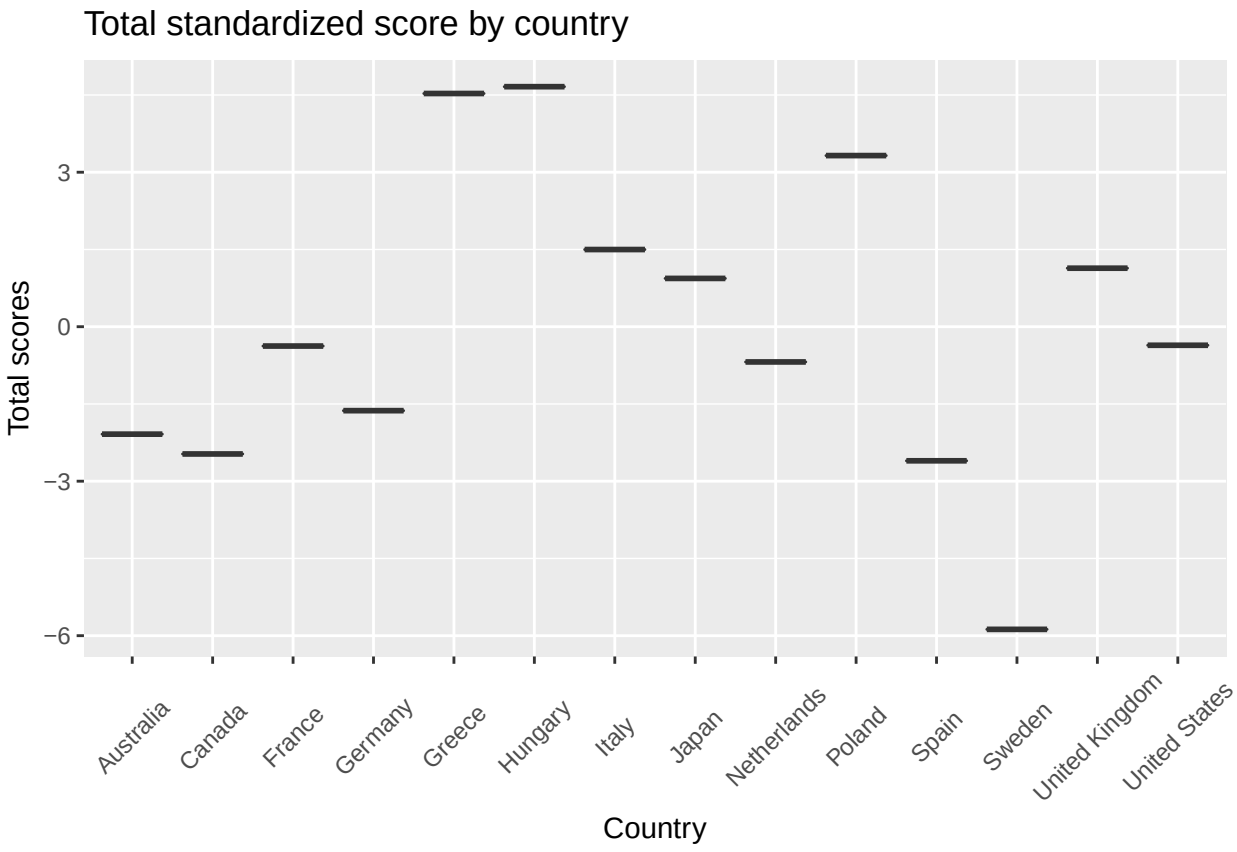
Erik Parker

February 14, 2017

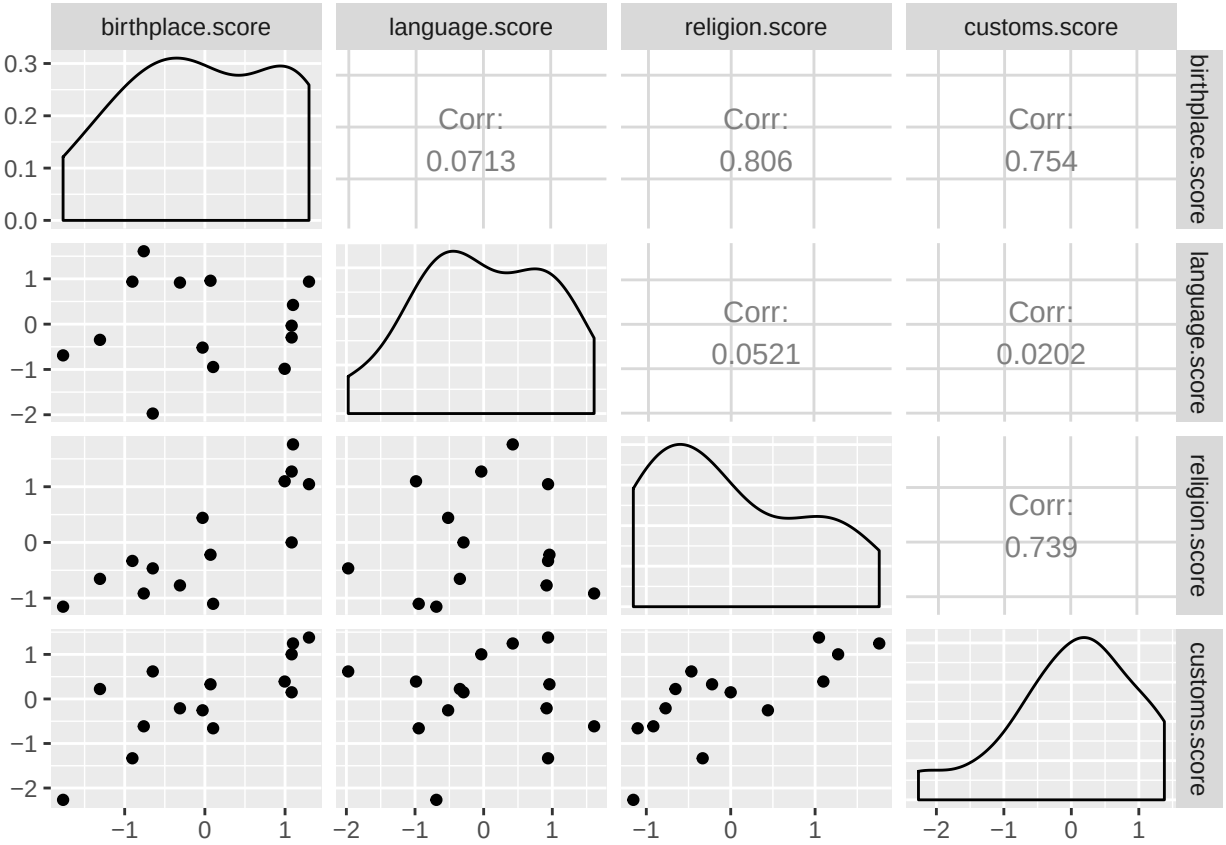
1. Univariate analysis. Find the total score for each country (adding up the standardized scores for all four questions) and

display the results in a table or graph.

```
##           Country total.scores
## 1   United States   -0.3601177
## 2          Canada   -2.4712147
## 3          France   -0.3751138
## 4         Germany   -1.6302856
## 5          Greece    4.5290474
## 6         Hungary    4.6614464
## 7           Italy    1.4988112
## 8     Netherlands   -0.6836998
## 9          Poland    3.3234405
## 10          Spain   -2.6038429
## 11         Sweden   -5.8769892
## 12 United Kingdom    1.1365780
## 13      Australia   -2.0860975
## 14          Japan    0.9380378
```


Total standardized score by country
Total scores
3
0
−3
−6
Australia
Canada
France
Germany
Greece
Hungary
Italy
Japan
Netherlands
Poland
Spain
Sweden
United Kingdom
United States
Country

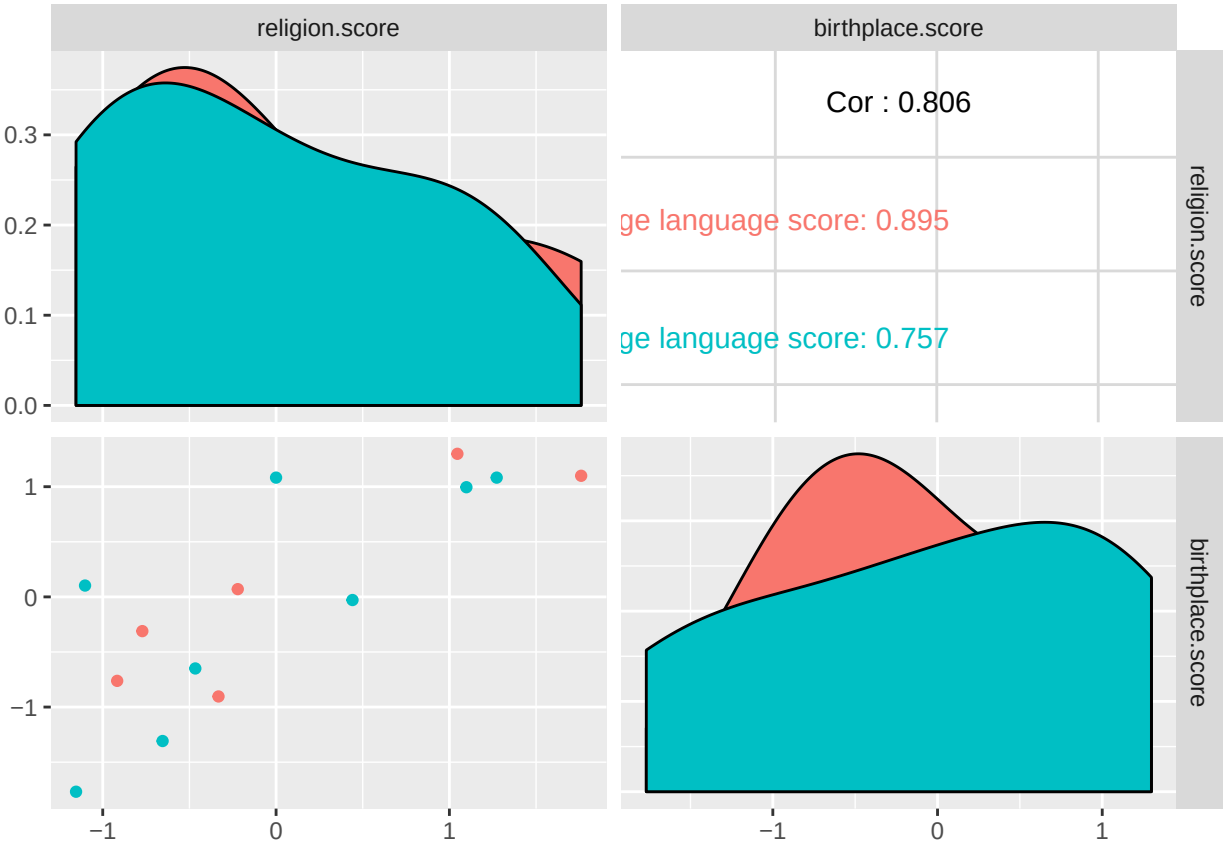
2. Bivariate analysis. Use ggpairs() in the GGally library to create a scatterplot matrix. There should be 4-choose-2 = 6 pairs of variables plotted. Identify which of these six pairs are strongly related, and which are weakly related.



From the pairwise correlation values shown in the plot above, it is clear that religion and birthplace are the most strongly related national identity variables across all countries, with birthplace and customs falling into a close second, and religion and customs in third. Conversely, language and customs, language and birthplace, and language and religion are quite weakly related. Meaning that in general: customs, religion, and birthplace are all closely correlated with one another.

Interstingly (given American religious right political talking points), language doesn't seem to be strongly related to any of the other metrics of national identity (vocal minorities strike again?).

3. Trivariate analysis. You should find that three of the variables are quite strongly related, while the other variable is more weakly related. Draw a scatterplot of the two most strongly correlated variables. Then color the points according to the weakly related variable (e.g. make the points where the weakly related variable is above average one color, and the weakly related variable is below average another color.) What does this tell you that was not obvious from the bivariate analysis?



This plot seems to be saying that the high correlation between birthplace and religion is complicated by the fact that respondants in countries which placed a higher than average importance on language show a much higher overall correlation between birthplace and religion than did respondants who placed a lower than average importance on language. That is, the overall correlation between religion and birthplace for all countries of 0.806 would be quite a bit higher (0.895) if we only focused our analysis countries that place a higher emphasis on the importance of language, or lower (0.757) if we focused on the other countries. So in conclusion, this extra trivariate analysis helps us see more of the underlying detail of the patterns first uncovered by the bivariate analysis.

‘‘‘