Students do NOT need to show R code in their submissions. Code is shown below just for informational purposes.

Exercise 1: Using the wide form GDP data, identify which countries are the outliers between 1971 and 1985. State clearly which they are. (*hint:* there are 3 which meet our criteria.)

```
> outliers = subset(gdp.wide, gdp.wide$'1980' > 15000)
> print(outliers$Country)
```

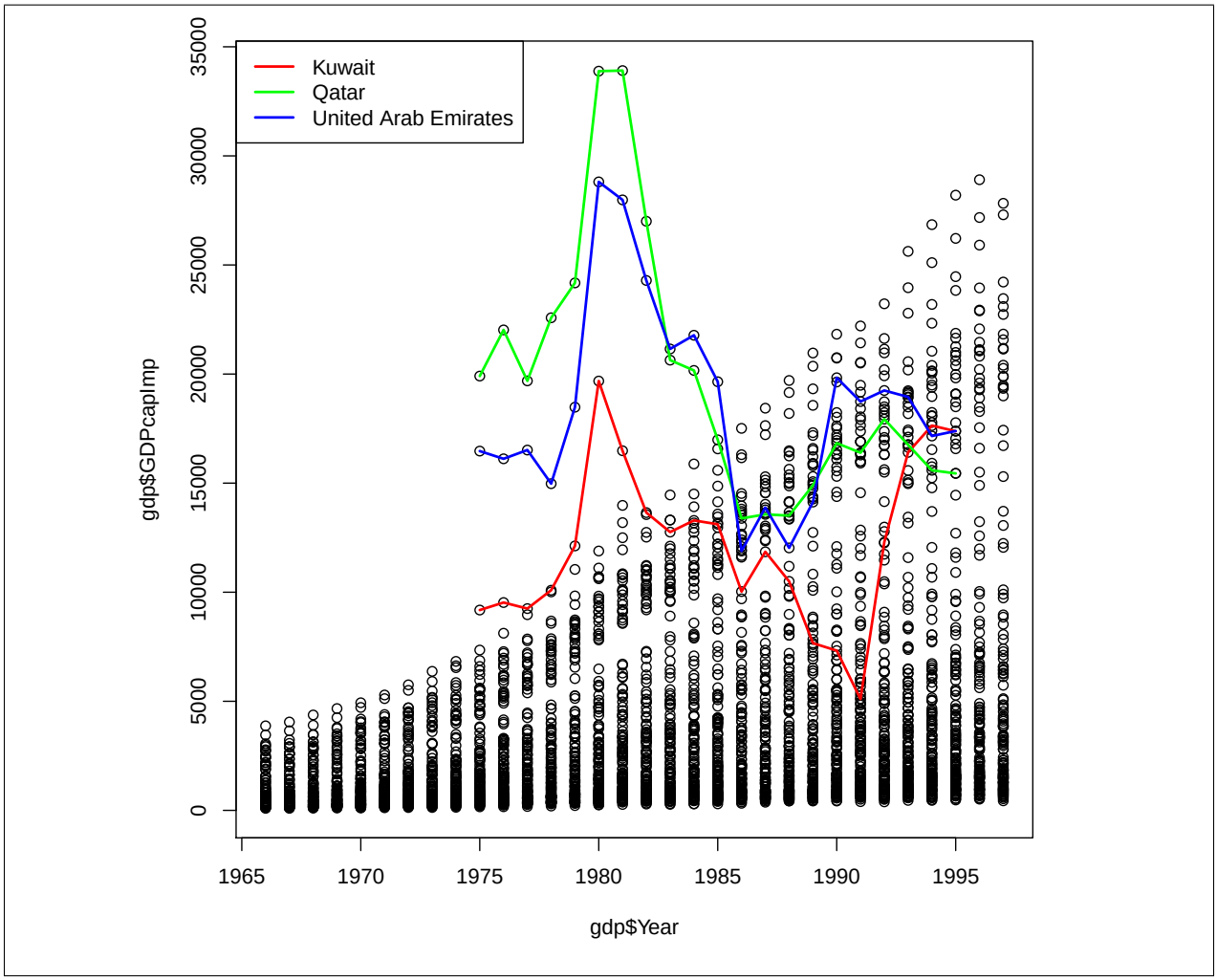
```
[1] "Kuwait"                "Qatar"
[3] "United Arab Emirates"
```

Exercise 2: Recreate the scatterplot at the top of this assignment. Add to this plot lines for the outlier countries you've identified. Be sure to color the lines so that we can see which country is which, and add a legend (*hint:* refer to lecture 6 notes, slides 12-18).

Dock points if legend missing or things obviously wrong with plot. Title and axis labels are optional. Do NOT dock for legend covering parts of graph – some issues with RStudio and legend scaling.

```
> xs = 1966:1997
> nrows = dim(outliers)[1]
> Cols = rainbow(nrows)
> plot(gdp$Year, gdp$GDPcapImp)
> for(i in 1:nrows) lines(xs, outliers[i, 2:33], col=Cols[i], lwd=2, xlab='year', ylab='GDP
> legend('topleft', outliers$Country, col=Cols, lwd=rep(2,nrows))
```

(plot on next page)


Kuwait
Qatar
United Arab Emirates
gdp$GDPcapImp
0
5000
10000
15000
20000
25000
30000
35000
1965
1970
1975
1980
1985
1990
1995
gdp$Year

Exercise 3: Answer using between 500 and 1000 words total:

- Briefly describe the pattern that you see.
- What do these countries share in common?
- Given the time period in which we see these patterns, what do you think might be the cause of the pattern?
- What information would you need to see whether your ideas are correct?

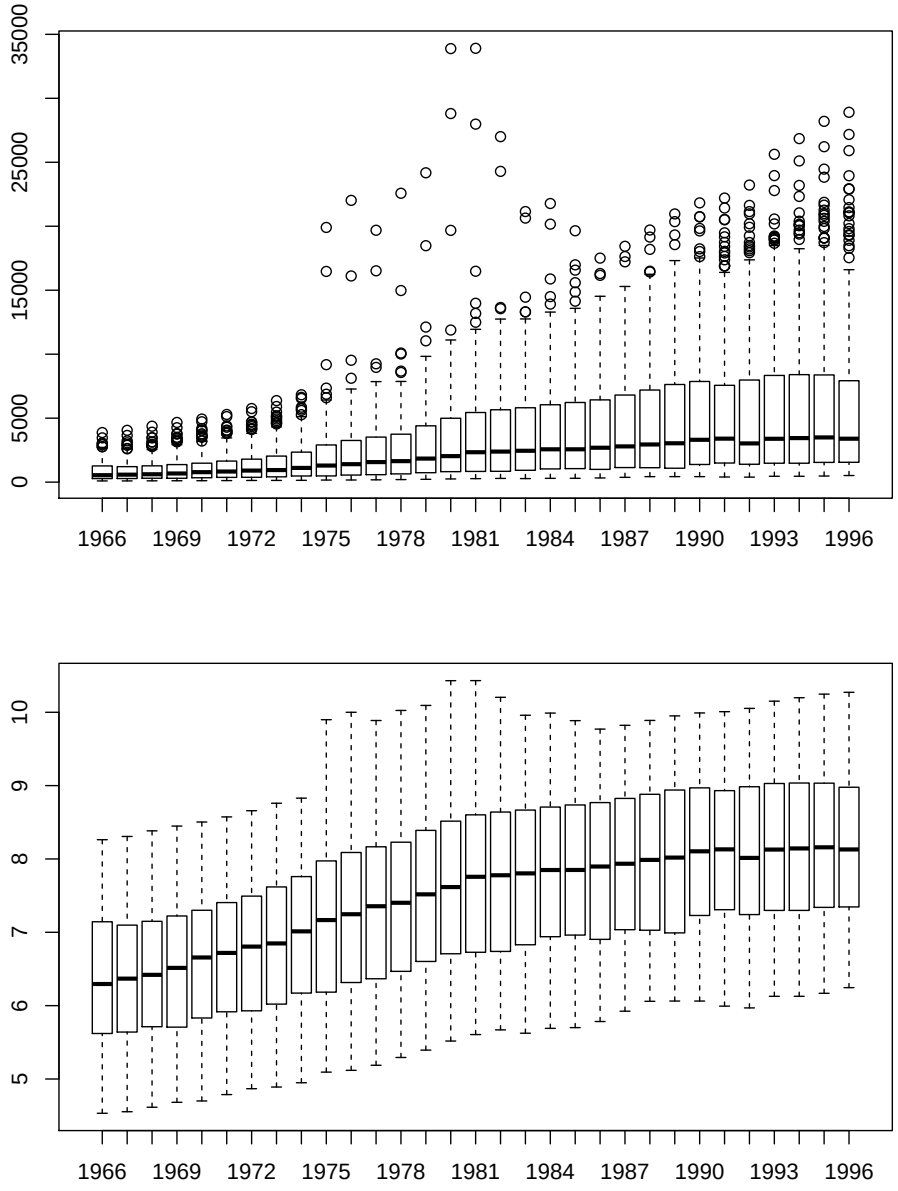
You'll need to read about these countries to answer this question. Remember to cite your sources. Given the casual nature of the question, Wikipedia is an acceptable source.

Any reasonable response that answers the above questions and meets the word limits is acceptable. Dock points for nonsensical statements, if any of the questions above are not addressed, not meeting word count reqs, and for improper citation.

Exercise 4: Create a side-by-side boxplot of (a) the raw GDP data, and (b) the log of the GDP data. In less than 500 words, describe, then compare and contrast the patterns you see in the two plots.

```
> par(mfrow=c(2,1))
> boxplot(gdp.wide[,2:32])
> boxplot(log(gdp.wide[,2:32]))
```

```
> par(mfrow=c(2,1))
> boxplot(gdp.wide[,2:32])
> boxplot(log(gdp.wide[,2:32]))
```



Dock for word count, talking about obviously erroneous stuff, not correctly creating the plot.

Exercise 5: Answer using between 250 and 750 words total:

- What feature of the log scale accounts for the differences in the way these two plots look? (*hint:* refer to lecture 5 notes, slides 7-17)
- What do these plots imply about change in GDP/capita over time?

For full credit, must state something approximating the following: raw boxplots suggest divergence between rich and poor countries. However, the log boxplots show that if we think about percent growth rather than magnitude, countries across the spectrum of wealth have done comparably. Further, must exhibit at least moderate understanding of what the log function does to right skewed data, by stating that it decompresses lower numbers, compresses higher numbers, and shows income to be log-normal.