Exercise 1:

1. Null and alternate hypotheses

   null the gap between parent populations is NO MORE than 20% big

   alternate the gap between parent populations is AT LEAST 20% big

2. Steps to test

   (a) Assume 1960 turnout normal by region

   (b) calculate sample mean, SD and size

   (c) run 2-tailed t-test with params from previous step and $\Delta = 20$

   (d) accept or reject based on p-val

Exercise 2:

1. Something similar to the below (hardcoded numbers is fine)

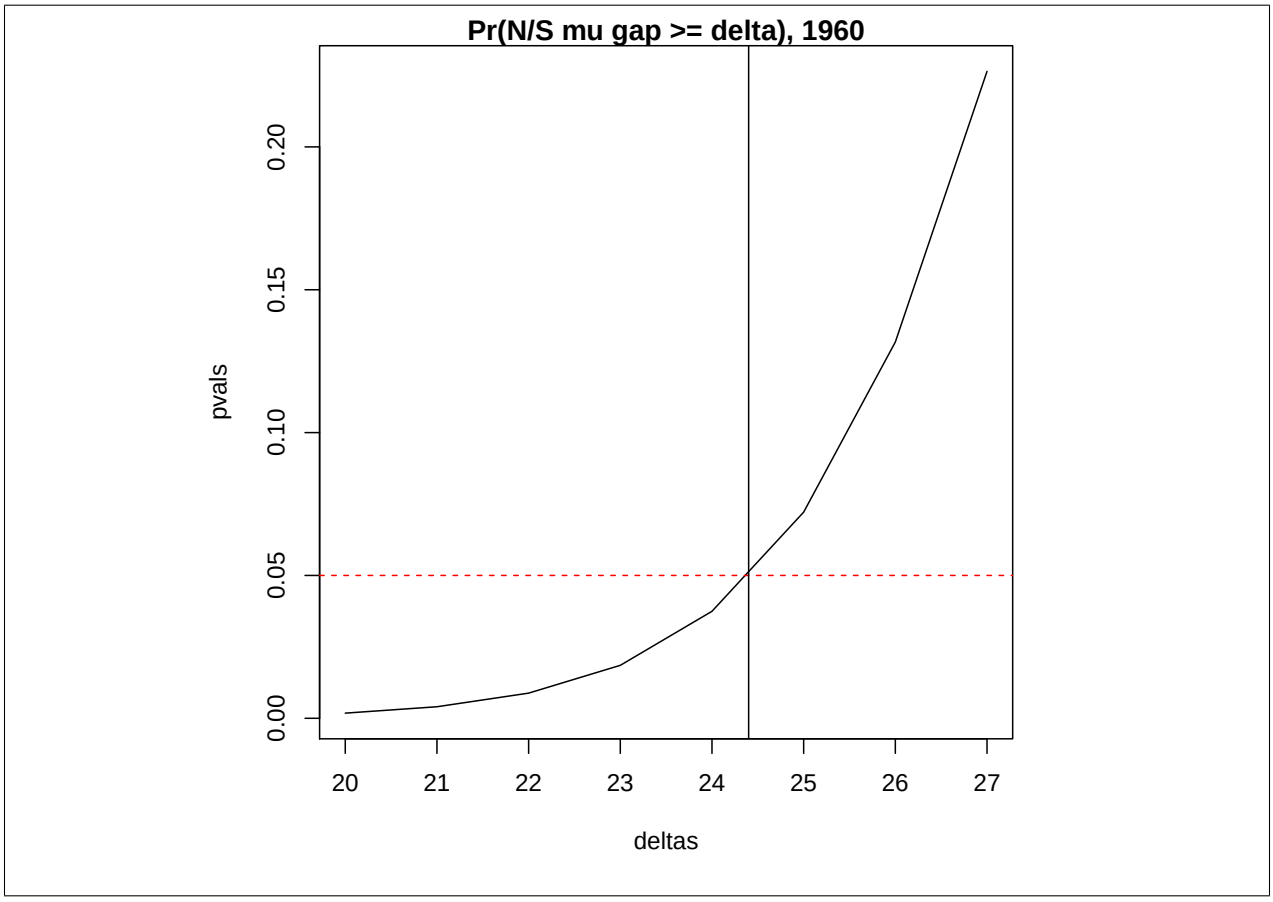
```
> s60 = turnout$p1960[turnout$deepsouth==1]
> n60 = na.omit( turnout$p1960[turnout$deepsouth==0] )
> t.test2( mean(n60), sd(n60), length(n60),
+                 mean(s60), sd(s60), length(s60), 20)

$pval
[1] 0.001852195

$T
[1] 3.442005

$nu
[1] 27.67715

>
> ## hardcoded
> #t.test2( 69.69487,  8.48465, 39, 38.78182, 9.50113, 11, 20)
```

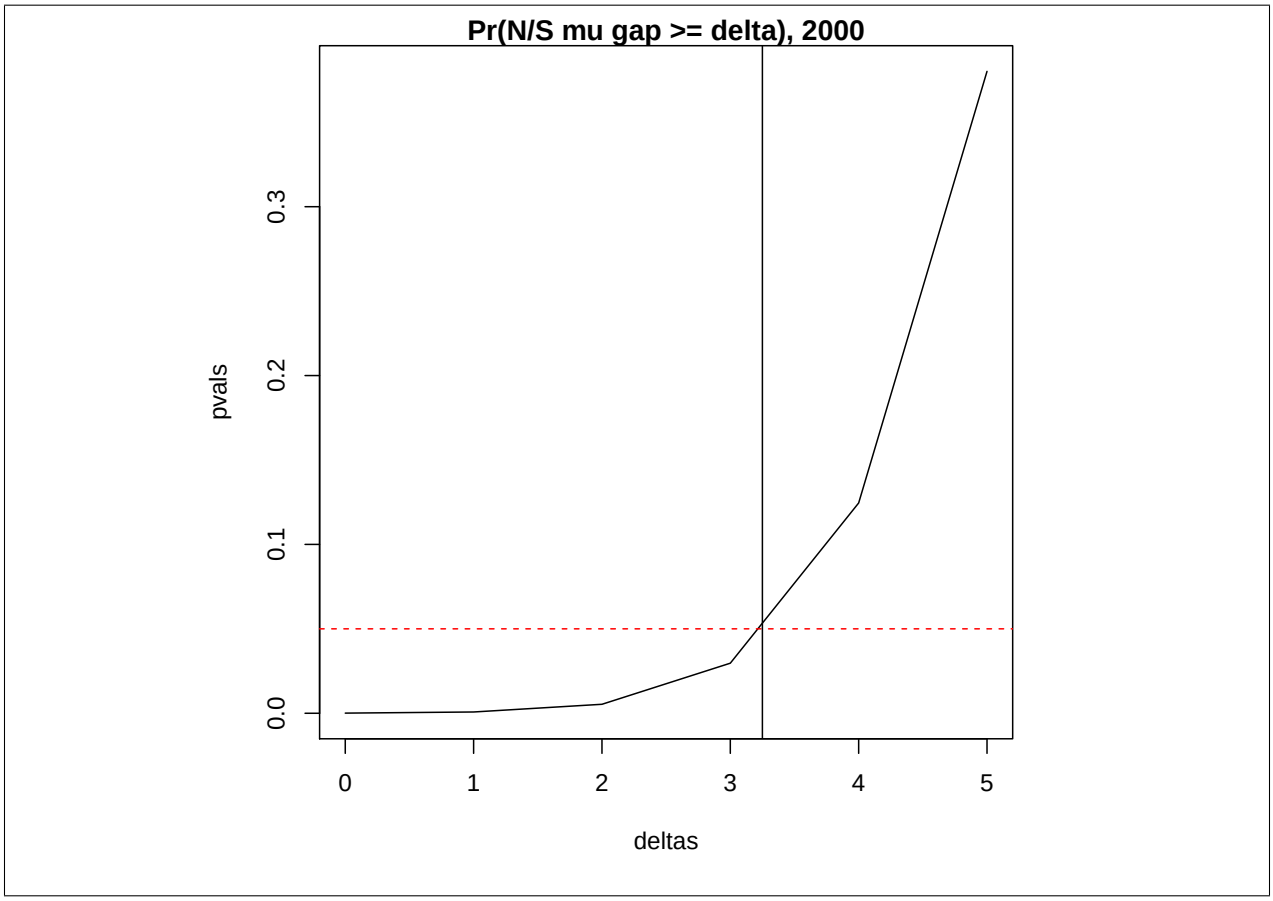
2. yes we can reject since < 0.05.

Exercise 3:

1. 1960: $\Delta \approx 24.4$ is max distance between the population means of South and North that is statistically significant.
2. 2000: $\Delta \approx 3.25$

See below plots for reasoning.





Exercise 4: Observation of decreasing gap is supported since the maximum gap between parent populations in 1960 is very much smaller than that in 2000. If repeated for interceding years, we'd probably see Δ_{max} from exercise 4 decrease across years. Additionally, grade based on whether there is anything incorrect stated, and whether word limit (250) is met.

Exercise 5: 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.)

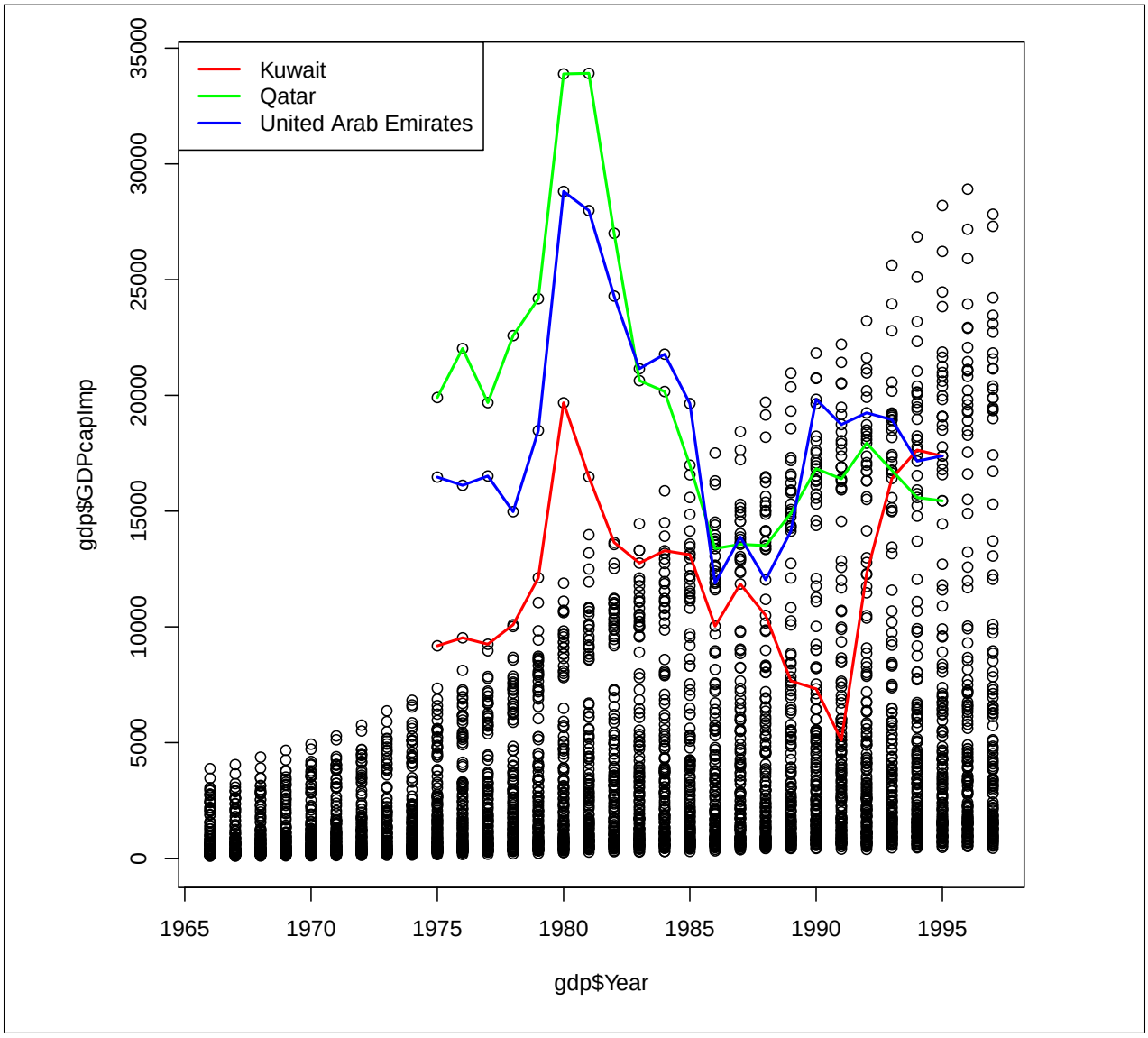
```
> outliers = subset(gdp.wide, gdp.wide$'1980' > 15000)
> print(outliers$Country)

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

Exercise 6: 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

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/cap')
> legend('topleft', outliers$Country, col=Cols, lwd=rep(2,nrows))
```

(plot on next page)


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

Exercise 7: 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.