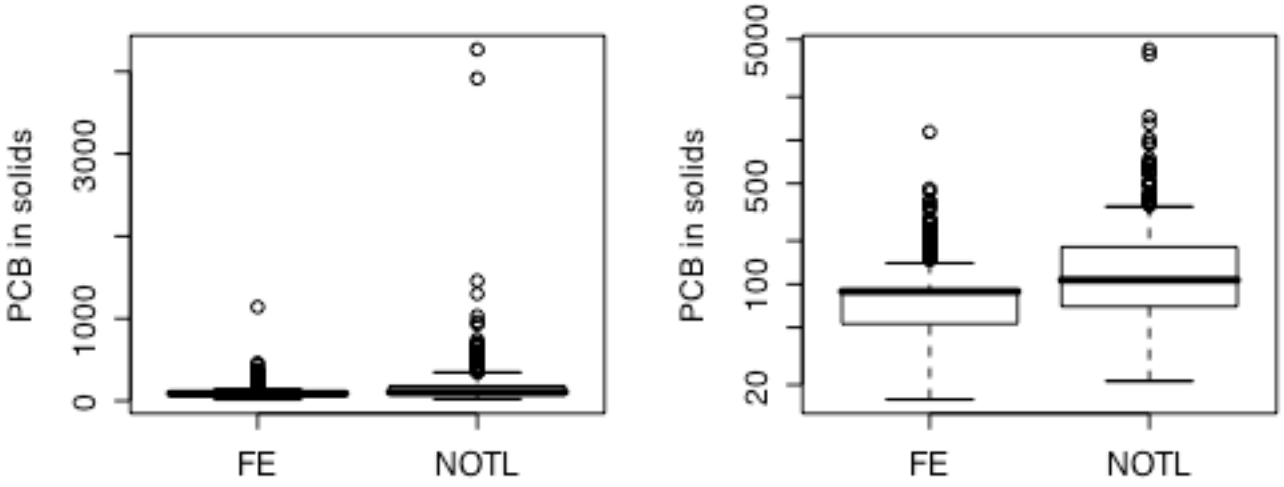
STATISTICS 3N03/3J04
Test #1 Solutions

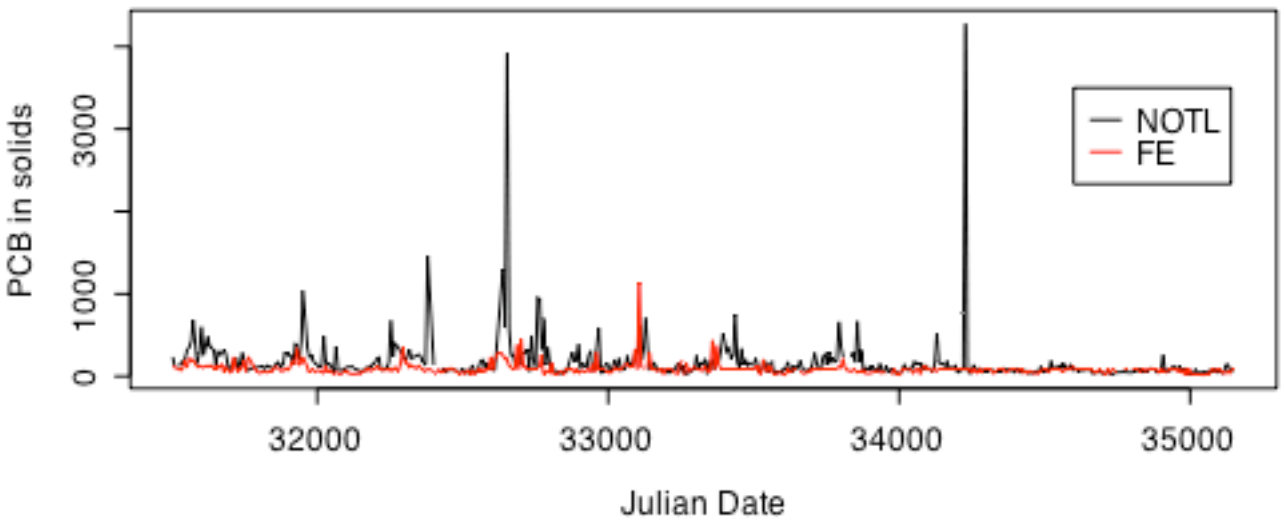
Question 1 [25 marks]

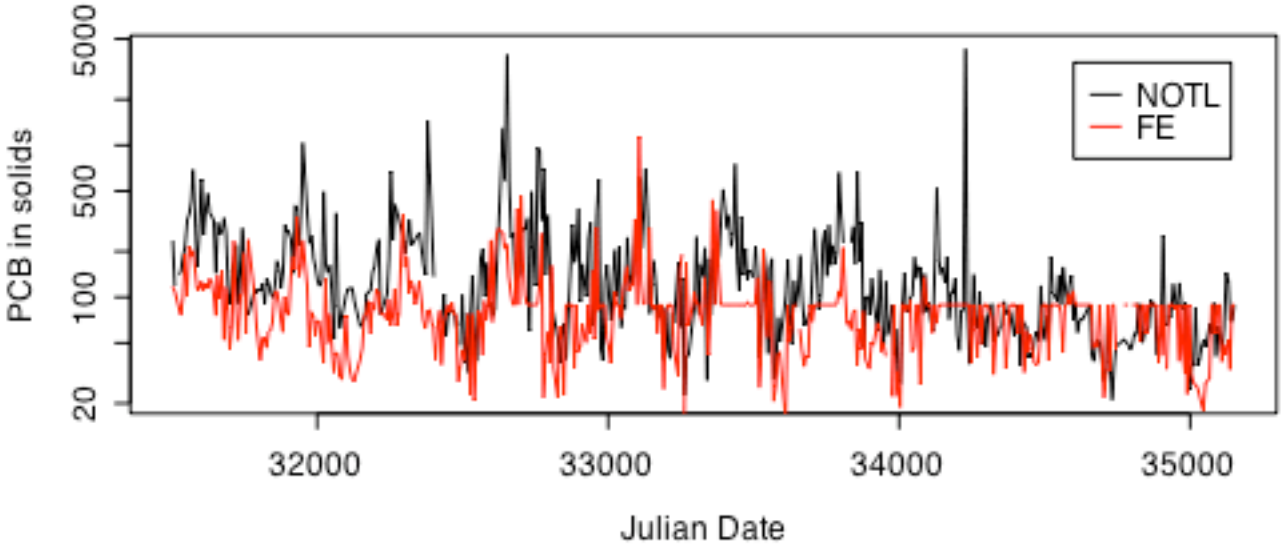
(a) When the time series data are aggregated over the entire time period and displayed in box plots, the median and quartiles are higher at Niagara on the Lake than at Fort Erie. Niagara on the Lake has more extreme high readings than Fort Erie.




```
> boxplot(PCB.SOL~STATION, niagara, ylab="PCB in solids")
> boxplot(PCB.SOL~STATION, niagara, ylab="PCB in solids",log="y")
```

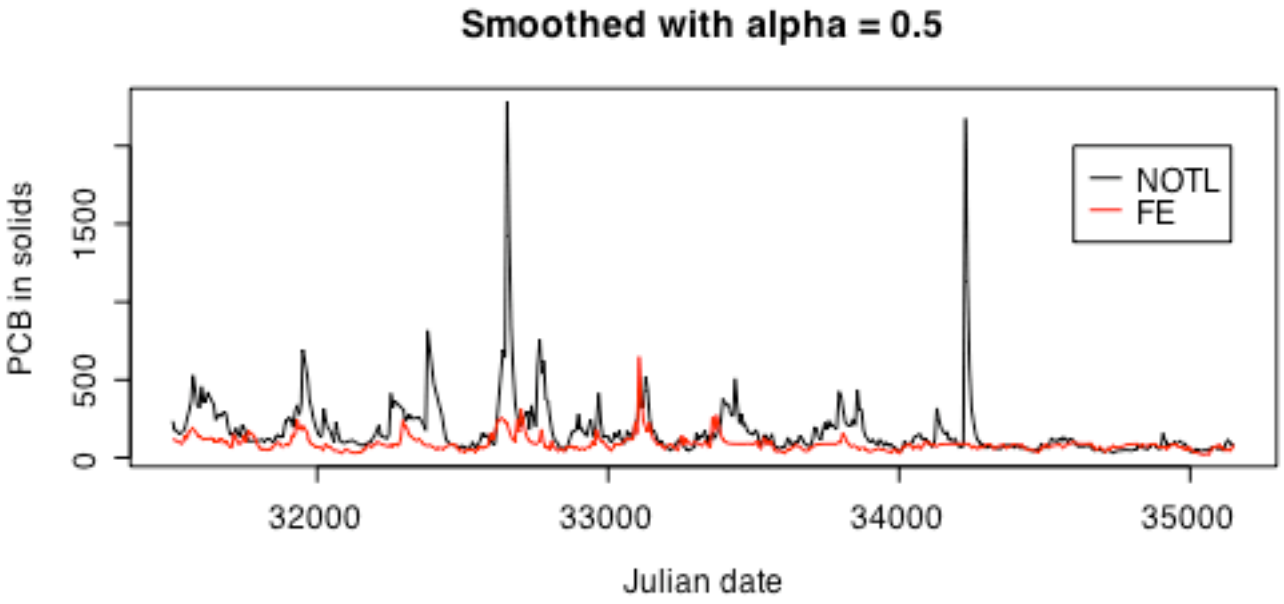
(b) Observed as a time series, pollution from PCB in Solids at Niagara on the Lake on any given day is almost always higher than at Fort Erie and has more extreme high peaks.

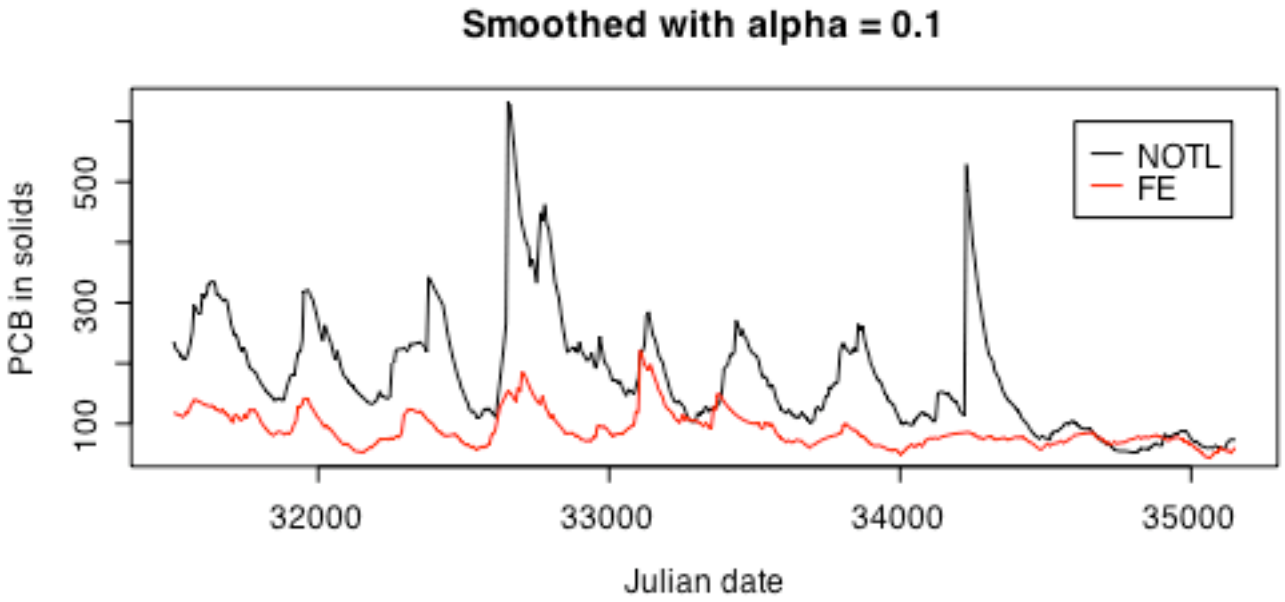





```
> plot(PCB.SOL~JULIAN.DATE, niagara[niagara$STATION=="NOTL",], xlab="Julian Date", ylab="PCB in
solids", type="l")
> lines(PCB.SOL~JULIAN.DATE, niagara[niagara$STATION=="FE",], col="red")
> legend(34600,3500,c("NOTL","FE"),lty=c(1,1),col=c(1,2))
> plot(PCB.SOL~JULIAN.DATE, niagara[niagara$STATION=="NOTL",], xlab="Julian Date", ylab="PCB in
solids", type="l", log="y")
> lines(PCB.SOL~JULIAN.DATE, niagara[niagara$STATION=="FE",], col="red")
> legend(34600,3500,c("NOTL","FE"),lty=c(1,1),col=c(1,2))
```

(c) Smoothing makes the difference between the series much clearer because it lowers the extreme peaks in the Niagara on the Lake data. The annual cycles are more evident. Especially with more smoothing (alpha = 0.1) it is clear that pollution is much higher at Niagara on the Lake up to the last few years when the series come together at the same level. This suggests that the amount of pollution entering the Niagara River has been reduced to an insignificant amount by the end of the study period.





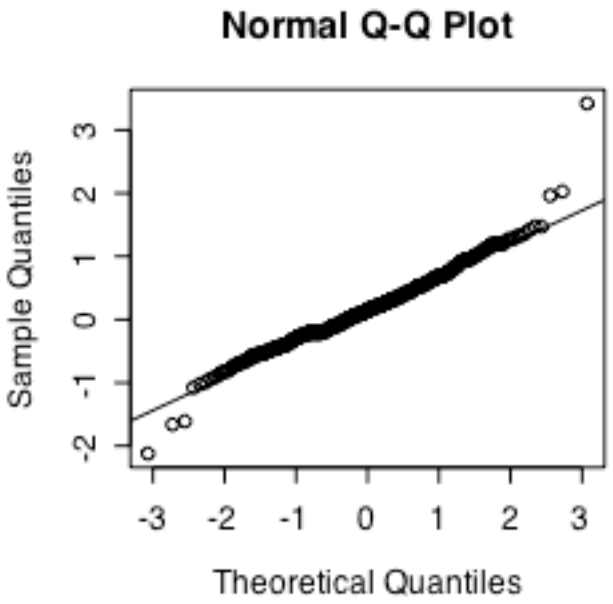

```
> expsmooth
function (x, alpha = 0.5)
{
    xs <- x
    for (i in 2:length(x)) {
        xs[i] <- alpha * x[i] + (1 - alpha) * xs[i - 1]
    }
    xs
}
> plot(expsmooth(PCB.SOL)~JULIAN.DATE, niagara[niagara$STATION=="NOTL" &
!is.na(niagara$PCB.SOL),],ylab="PCB in solids", xlab="Julian date", type="l")
> lines(expsmooth(PCB.SOL)~JULIAN.DATE, niagara[niagara$STATION=="FE" & !is.na(niagara$PCB.SOL),],
col="red")
> legend(34600,2000,c("NOTL","FE"),lty=c(1,1),col=c(1,2))
> title(main="Smoothed with alpha = 0.5")
> plot(expsmooth(PCB.SOL, 0.1)~JULIAN.DATE, niagara[niagara$STATION=="NOTL" &
!is.na(niagara$PCB.SOL),],ylab="PCB in solids", xlab="Julian date", type="l")
> lines(expsmooth(PCB.SOL, 0.1)~JULIAN.DATE, niagara[niagara$STATION=="FE" &
!is.na(niagara$PCB.SOL),], col="red")
> legend(34600,600,c("NOTL","FE"),lty=c(1,1),col=c(1,2))
> title(main="Smoothed with alpha = 0.1")
```

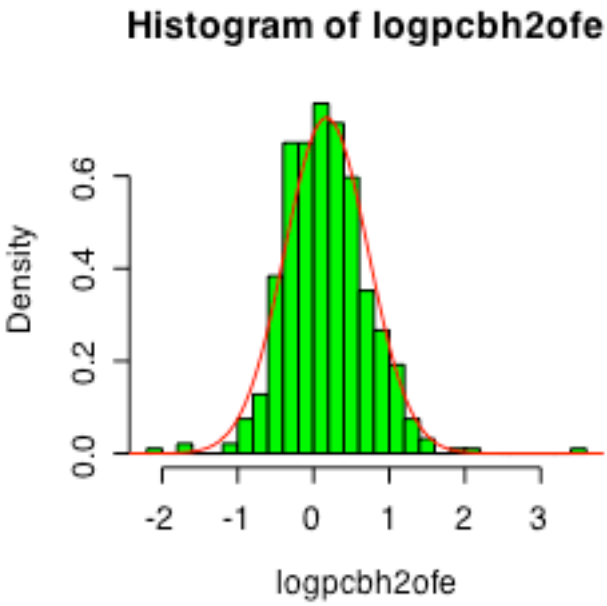
(d) Log transforming the unsmoothed time series reduces the visual impact of the extreme high peaks in the Niagara on the Lake data and shows the fluctuations at lower levels of pollution more clearly. It is interesting to see that there is an upper limit of detection in the Fort Erie series in the second decade. This means that pollution at Fort Erie will be underestimated but that will not change our conclusion because the concentration at Niagara on the Lake is no higher than the underestimated Fort Erie concentration.

The smoothed (alpha = 0.1) time series concentrations all lie within an order of magnitude, so a log transformation would not make a noticeable difference and I didn't try it. The smoothed time series is a very effective way to show that the amount of PCB in solids entering the Niagara River between Fort Erie and Niagara on the Lake was reduced to an insignificant amount by the final years of the study.

(e) Test #1a – PCB in Water at Fort Erie

The distribution of log of PCB concentration is very close to Normal, except for one outlier on the right tail.

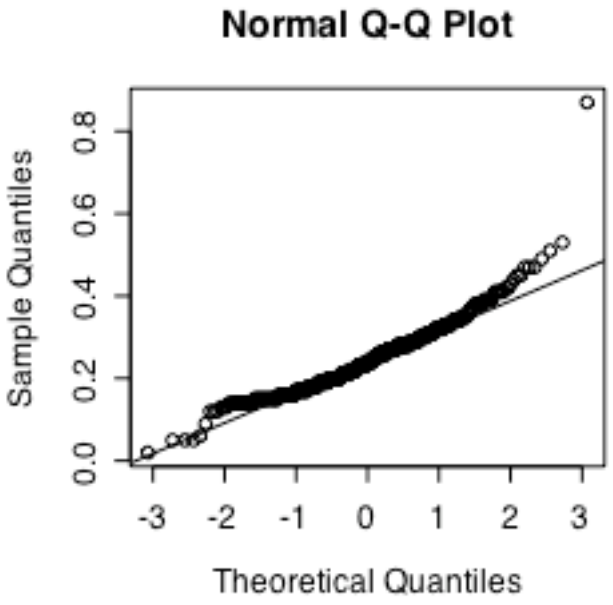


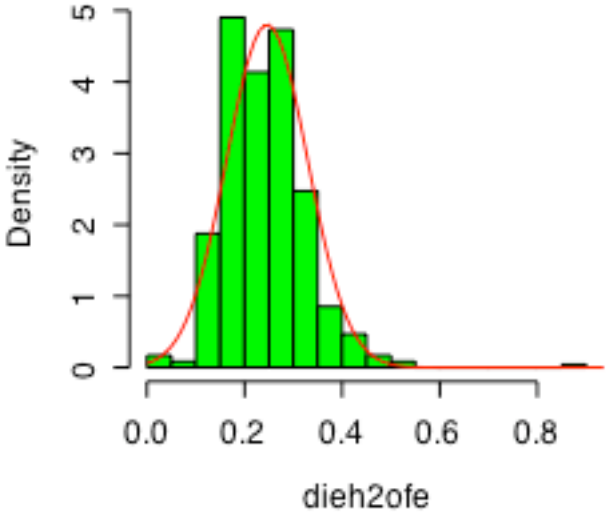


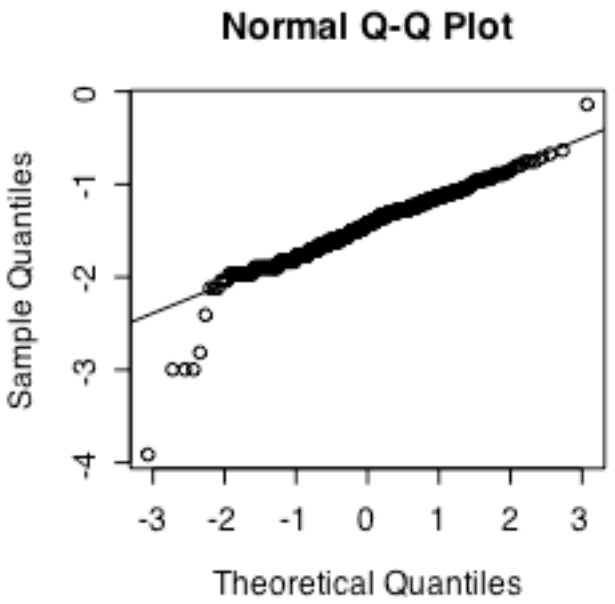

```
> logpcbh2ofe <- log(niagara$PCB.H2O[niagara$STATION=="FE"])
> qqnorm(logpcbh2ofe)
> qqline(logpcbh2ofe)
> hist(logpcbh2ofe, breaks=25, prob=T, col="green")
> lines(xgr, dnorm(xgr,mean(logpcbh2ofe, na.rm=T),sd(logpcbh2ofe, na.rm=T)), col="red")
```

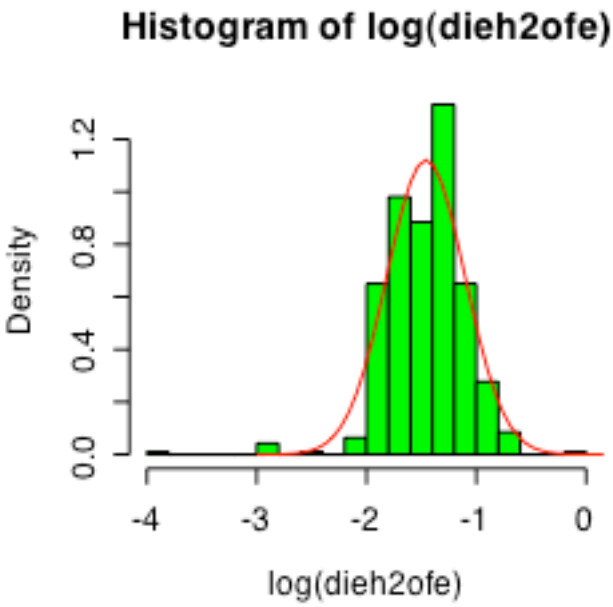
(e) Test #1b – Dieldrin in Water at Fort Erie

The distribution of Dieldrin concentration in water is close to Normal, but is slightly skewed to the right and has one outlier in the right tail. The distribution of the log of Dieldrin concentration in water is close to Normal, except for some outliers in the left tail.





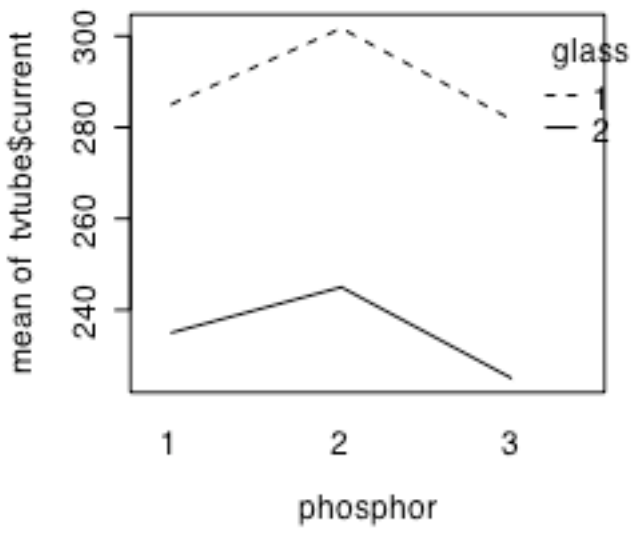


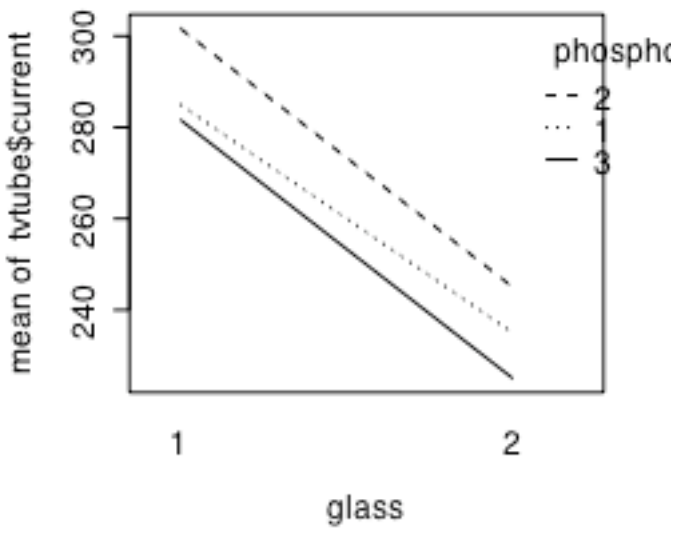


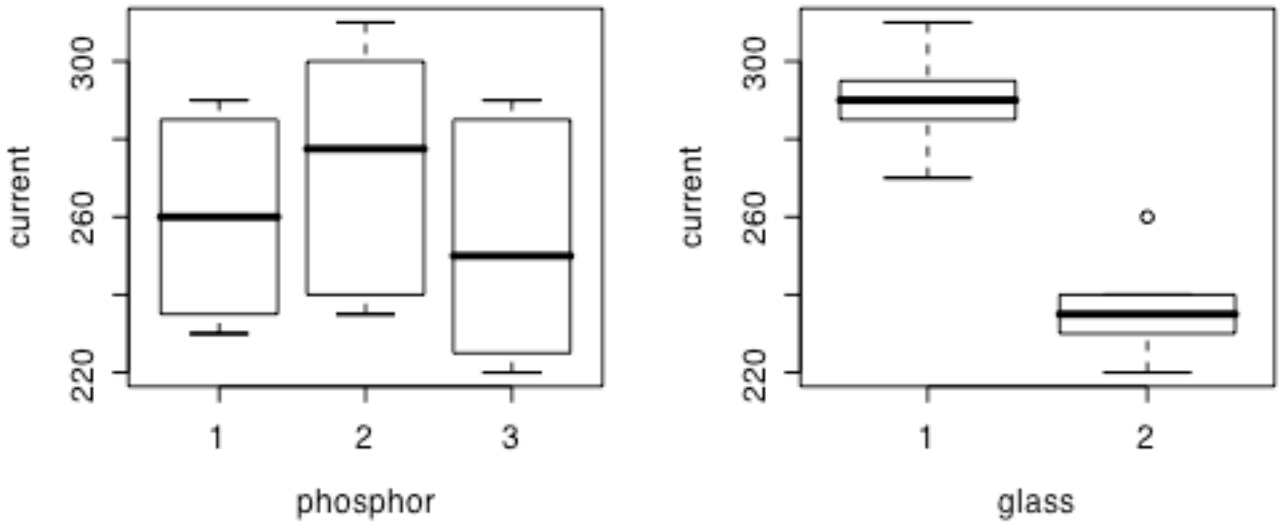

```
> dieh2ofe <- niagara$DIE.H2O[niagara$STATION=="FE"]
> qqnorm(dieh2ofe)
> qqline(dieh2ofe)
> hist(dieh2ofe, breaks=25, prob=T, col="green")
> xgr <- seq(0,1,length=101)
> lines(xgr, dnorm(xgr,mean(dieh2ofe, na.rm=T),sd(dieh2ofe, na.rm=T)), col="red")
> qqnorm(log(dieh2ofe))
> qqline(log(dieh2ofe))
> hist(log(dieh2ofe), breaks=25, prob=T, col="green")
> xgr <- seq(-3,4,length=101)
> lines(xgr, dnorm(xgr,mean(log(dieh2ofe), na.rm=T),sd(log(dieh2ofe), na.rm=T)), col="red")
```

Question 2 **[12 marks]**

The interaction plots are parallel, so there is no evidence of interaction between glass and phosphor. For any of the three phosphors, glass 2 requires the least current to achieve the same brightness and for either glass phosphor 3 requires the least current and phosphor 2 the greatest current to achieve the same brightness.







I avoided having to re-type the data by copying the table from the web page and pasting it into Word, editing so that the numbers were all separated by commas, and then copying and pasting the edited data into the R command below.

```
tvtube <- data.frame(current=c(280, 290, 285, 300, 310, 295, 270, 285, 290, 230, 235, 240, 260,
240, 235, 220, 225, 230), glass=factor(rep(1:2,rep(9,2))), phos=factor(rep(rep(1:3,rep(3,3)),2)))
> tvtube
   current glass phos
1      280     1    1
2      290     1    1
3      285     1    1
4      300     1    2
5      310     1    2
6      295     1    2
7      270     1    3
8      285     1    3
9      290     1    3
10     230     2    1
11     235     2    1
12     240     2    1
13     260     2    2
14     240     2    2
15     235     2    2
16     220     2    3
17     225     2    3
18     230     2    3
> interaction.plot(tvtube$phos,tvtube$glass,tvtube$current, trace.label="glass",xlab="phosphor")
> interaction.plot(tvtube$glass,tvtube$phos,tvtube$current, trace.label="phosphor",xlab="glass")
> boxplot(current~phos, tvtube, xlab="phosphor", ylab="current")
> boxplot(current~glass, tvtube, xlab="glass", ylab="current")
```

Question 3 **[3 marks]**

Give any three interesting facts about the person chosen. Refer to the links on the course web page for their biographies.

Bonus Marks: **[up to 5 marks for the quality of writing and presentation]**

FULL MARKS = 45