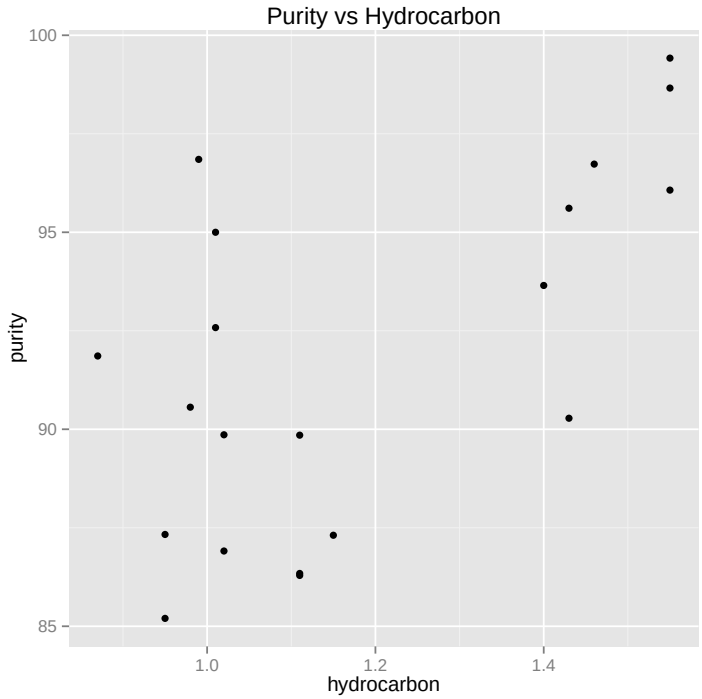
Question 2

A: Plot Purity vs Hydrocarbon. Discuss what you see in relation to the SLR assumptions.

```
> ## loading usefull packages
> library(knitr);library(xtable);library(ggplot2)
> ## read in data
> data<-read.table("A1_data.txt",sep=" ",header=T)
> ## rename the variables
> names(data)<-c("purity","hydrocarbon")
> ## Plot graph
> plot_1<-qplot(hydrocarbon,purity,data=data,main="Purity vs Hydrocarbon")
> print(plot_1)
```



B: Fit a linear model and plot the fitted line to (A). Explain the model

```
> fit1<-lm(purity~hydrocarbon,data=data)
> print(xtable(summary(fit1)))
```

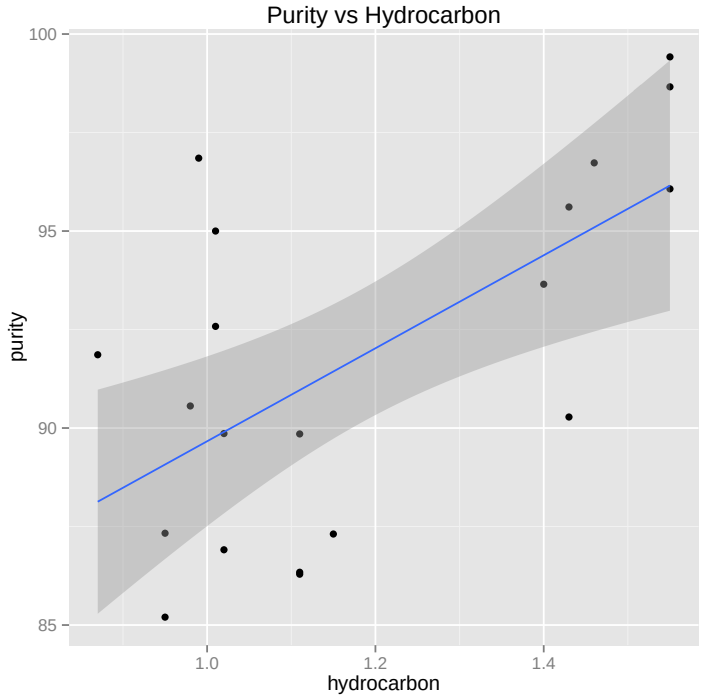
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	77.8633	4.1989	18.54	0.0000
hydrocarbon	11.8010	3.4851	3.39	0.0033

```
> print(xtable(anova(fit1)))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
hydrocarbon	1	148.31	148.31	11.47	0.0033
Residuals	18	232.83	12.94		

The model parameter.....

```
> plot_1<-plot_1+geom_smooth(method = "lm")
> print(plot_1)
```



C: Fit a 95 percent prediction and confidence interval for the purity level when the hydrocarbon percentage is equal to to 1.0. Explain the PI and CI.

```
> ## need to check
> confidence_interval<-predict(fit1,newdata=data.frame(hydrocarbon=1),
+                              interval=("confidence"),level=0.95)
> prediction_interval<-predict(fit1,newdata=data.frame(hydrocarbon=1),
+                              interval=("prediction"),level=0.95)
> print(confidence_interval)

       fit      lwr      upr
1 89.66431 87.51017 91.81845

> print(prediction_interval)

       fit      lwr      upr
1 89.66431 81.80716 97.52146
```

A confidence interval expresses uncertainty about the expected value of y-values at a given x. A prediction interval expresses uncertainty surrounding the predicted y-value of a single sampled point with that value of x.

D: Do the hypothesis test for B1=0 at 0.05 level of significance. Explain.

```
> print(xtable(summary(fit1)))
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	77.8633	4.1989	18.54	0.0000
hydrocarbon	11.8010	3.4851	3.39	0.0033

As you can see the table summary of coefficients automatically calculates the hypothesis test for me.