Find a published clinical trial on a biomedical problem of interest. Read and write a short report. Identify the key components, including the purpose of the study, main outcome variable, risk factors, data collection procedure, and key findings. Discuss: is the finding on association or causation? Is it possible to conduct an observational study

The source

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8201647/?report=classic#sec1

The key components

1. purpose of the study：

Searching for the factors that influence risk of hospital admission and vaccine effectiveness.

2. main outcome variable：

Every part of outcome variable is not same in this report. In some part researchers use the value of Hazard Ratios (HR) to identify the vaccine effect and in some part researchers use the value of Vaccine Effect to determine the effectiveness of a particular brand of vaccine. Thus, we think that the vaccine effectiveness of protection is outcome variable.

3. risk factors：

* different type of Covid-19: Alpha VOC and Delta VOC
* Age
* Gender
* Deprivation

4. data collection procedure：

* The data of S gene if from the analysis using ThermoFisher's TaqPath RT-PCR.

* The researchers through EAVE II to collect the demographic profile of COVID-19 patients, the risk of hospital admission for COVID-19.

EAVE II is a Scotland-wide COVID-19 surveillance platform that has been used to track and forecast the epidemiology of COVID-19, inform risk stratification, and investigate vaccine effectiveness and safety.1, 2, 3, 4 It comprises national health-care datasets on 5·4 million people (about 99% of the Scottish population) linked through Scotland's unique Community Health Index number.

5. key findings：

*S gene-positive cases were associated with an increased risk of COVID-19 hospital admission

*COVID-19 relevant comorbidities increased the risk of COVID-19 hospital admission

* (vaccine effect) with an interaction test p-value is 0.19. it suggest that there was no evidence of a differential vaccine effect on hospital admissions among those first testing positive

*(different brand of vaccine)

**for Pfizer vaccine

Compared to those unvaccinated, at least 14 days after the second dose, Pfizer–BioNTech vaccine offered good protection for S-gene negative disease (92% (95% CI 90–93)) and S-gene positive disease (79% (95% 75–82)).

** for AstraZeneca vaccine

Compared to those unvaccinated, at least 14 days after the second dose, Oxford–AstraZeneca vaccine also offered protection for S-gene negative disease (73% (95% CI 66–78)) and for S-gene positive disease(60% (95% CI 53–66))

** These can conclude that compare to AstraZeneca vaccine, Pfizer vaccine have a better performance

**Delta VOC mostly appeared in young affluent populations.

**The risk of COVID-19 hospital admission was doubled with Delta VOC compared to Alpha VOC.

Discuss

1.Is the finding on association or causation？

*This finding is on association. Because the study collected and analyzed the data from EAVE II but it didn’t carry out any confirmation experiment. Cohort analyses were performed using the EAVE II platform to describe a demographic profile of COVID-19 patients, investigate the risk of admission to hospital for COVID-19, and estimate the effectiveness of vaccines in preventing hospitalizations for COVID-19 in S-gene positive cases. The analysis was based on all individuals who underwent SARS-COV-2 PCR testing during the study period and compared the proportion of individuals who were vaccinated at the time of swab testing to those who were not vaccinated at the time of testing, adjusting for demographic and temporal covariables. The effect of the vaccine on hospitalization for COVID-19 can be estimated based on this study. In S-gene negative cases, the effect of vaccination (at least 28 days after the first or second dose) was to reduce the risk of hospitalization compared with non-vaccination

2. Is it possible to conduct a controlled experiment to address the same biomedical question?

Experimental Design:

In this randomized, double-blind, placebo-controlled field experiment, 1800 novel coronavirus with negative nucleic acid test, positive S gene, no allergy history, and no adverse reactions to the two vaccines used in the experiment were selected as the study subjects.The vaccines were given out randomly, and the volunteers were unaware of the type of vaccine they were receiving and whether it was placebo.

Participants were divided into two groups, each with 900 participants. Each participant was then divided into three groups (n=300) to receive the Pfizer vaccine, the Astrazeneca vaccine, and the placebo. After 7 days, 14 days and 28 days, the infection rates of alpha virus and Delta were detected, and the risk factors of hospitalization were calculated.

	Injection type	number	The prevalence of Alaph virus infection			The prevalence of Delta virus infection			Hospitalization risk factor
			On the 7th day	On the 14th day	On the 28th day	On the 7th day	On the 14th day	On the 28th day	
Male	Pfizer vaccine	300							
	Astrazeneca vaccine	300							
	Placebo	300							
Female	Pfizer vaccine	300							
	Astrazeneca vaccine	300							
	Placebo	300							

Homework2

Xu Guo

2022/7/28

R work

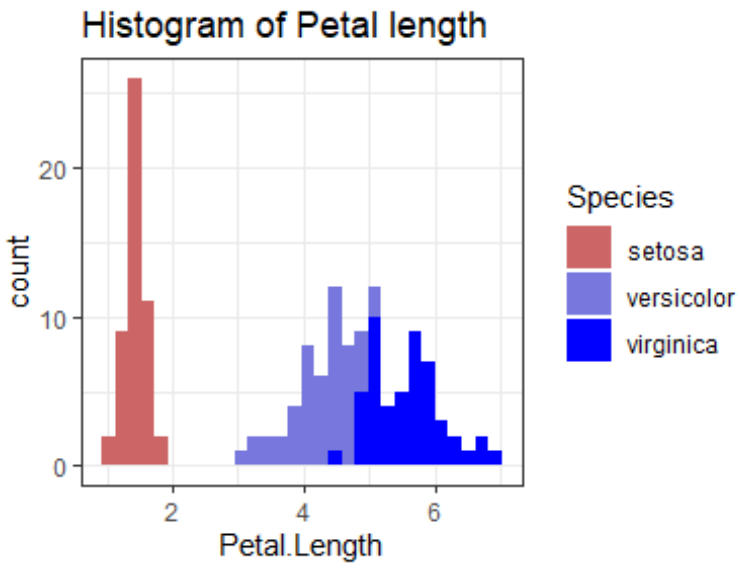
Using the built-in dataset “iris” to conduct EDA. ## Question1: Draw a density plot of Sepal.Width and a histogram of Petal.Length, remember to add suitable titles and choose whatever colour you like.

Data Presentation：

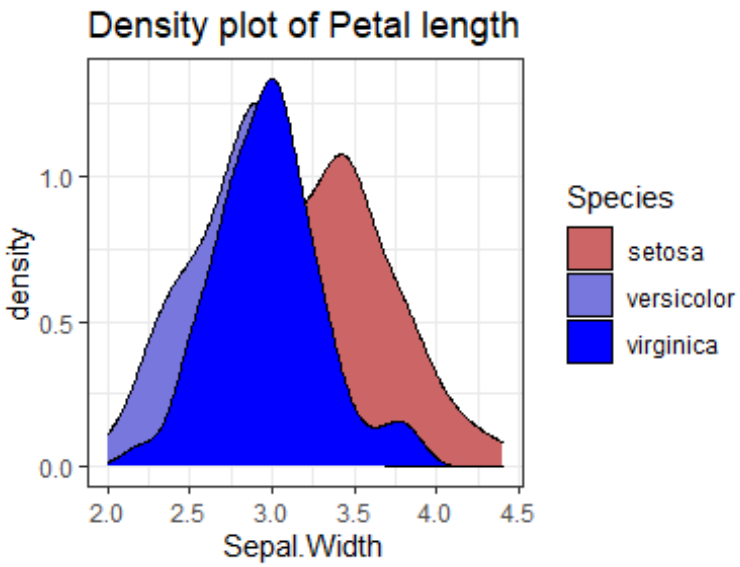
```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

R Codes:

```
library(ggplot2)
library(gcookbook)
ggplot(iris, aes(x = Petal.Length, fill = Species)) +
geom_histogram()+theme_bw()+labs(title = ("Histogram of Petal length
"))+scale_fill_manual(values=c("#CC6666","#7777DD","blue"))
```



```
ggplot(iris, aes(x = Sepal.Width, fill = Species))+ geom_density( )+t
heme_bw()+labs(title = ("Density plot of Petal length"))+scale_fill_m
anual(values=c("#CC6666","#7777DD","blue"))
```



Queston2:

Compute the summary statistics of Petal.Length (mean, median, sd, quantiles, etc).

R Code:

```
summary(iris$Petal.Length)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.600   4.350   3.758   5.100   6.900
```

Homework3

Xu Guo

2022/8/4

R work

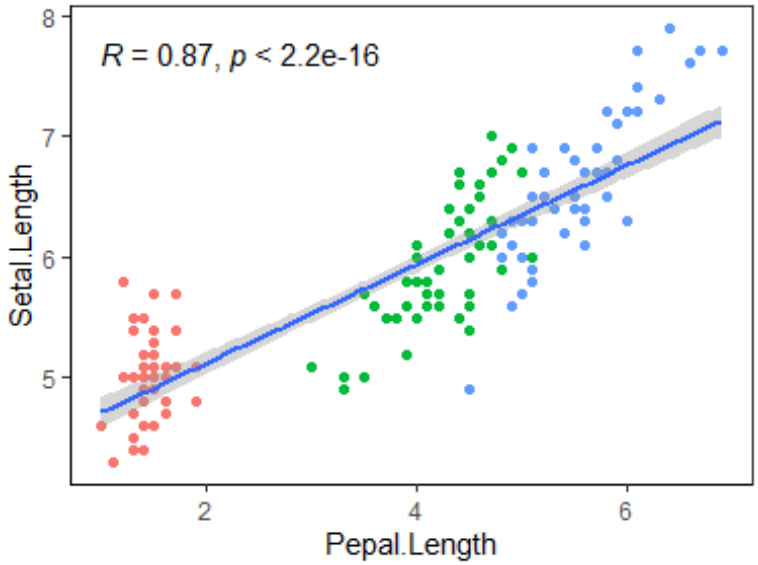
Using the built-in dataset “iris” to conduct EDA. ## Question1: Draw a scatterplot or a line plot of Petal.Length against Sepal.Length to study their association. Comment on the relationship.

Data Presentation：

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

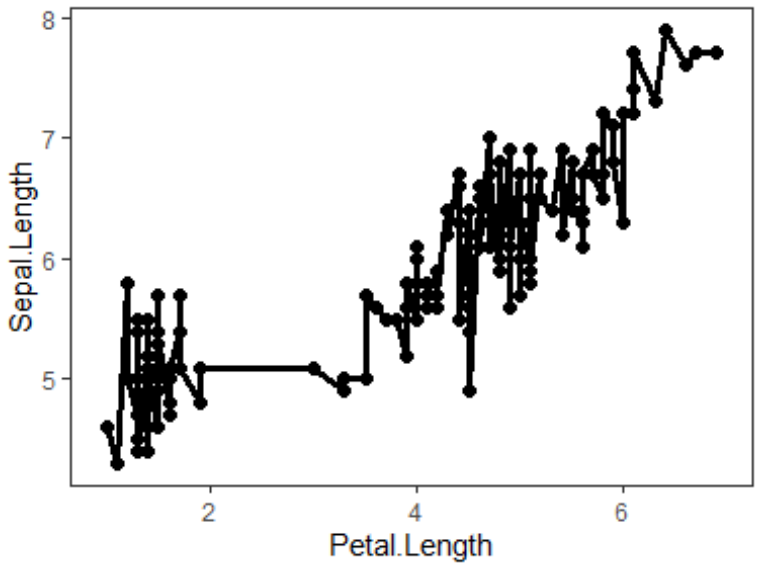
The scatterplot:

```
library(ggplot2)
library(tidyverse)
library(gcookbook)
library(ggpubr)
ggplot(iris, aes(iris$Petal.Length, iris$Sepal.Length)) +
  xlab("Pepal.Length")+ylab("Setal.Length")+
  geom_point(aes(colour=Species))+
  geom_smooth(method=lm) +
  theme_bw()+theme(panel.grid.major =element_blank(), panel.grid.mino
r = element_blank(),panel.background = element_blank(),axis.line = el
ement_line(colour = "black"))+stat_cor(aes(x =iris$Sepal.Length, y =i
ris$Petal.Length))+guides(size="none",colour = "none")
```



The line plot:

```
ggplot(iris,aes(Petal.Length,Sepal.Length))+
  geom_point(size=2)+geom_line(position = position_dodge(0.1),cex=1.3)
+theme(legend.title = element_blank())+theme_test()
```



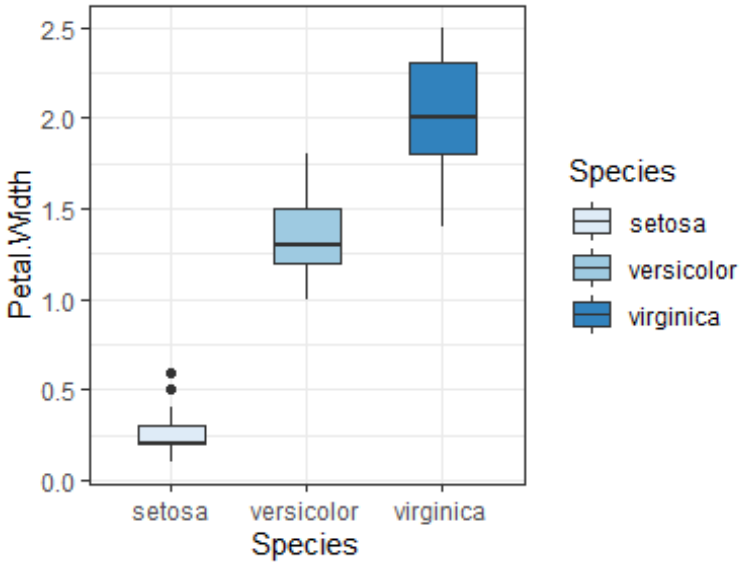
There is a strong correlation between petal length and sepal length

Queston2:

Use a boxplot to describe the distribution of Petal.Width against Species.

The boxplot:

```
ggplot(iris,aes(Species,Petal.Width,fill=Species))+geom_boxplot(widt
h=0.5)+theme_bw()+scale_fill_brewer()
```



Queston3: Come up with a different way of constructing new variables, and conduct some EDA based on the new variable you construct.

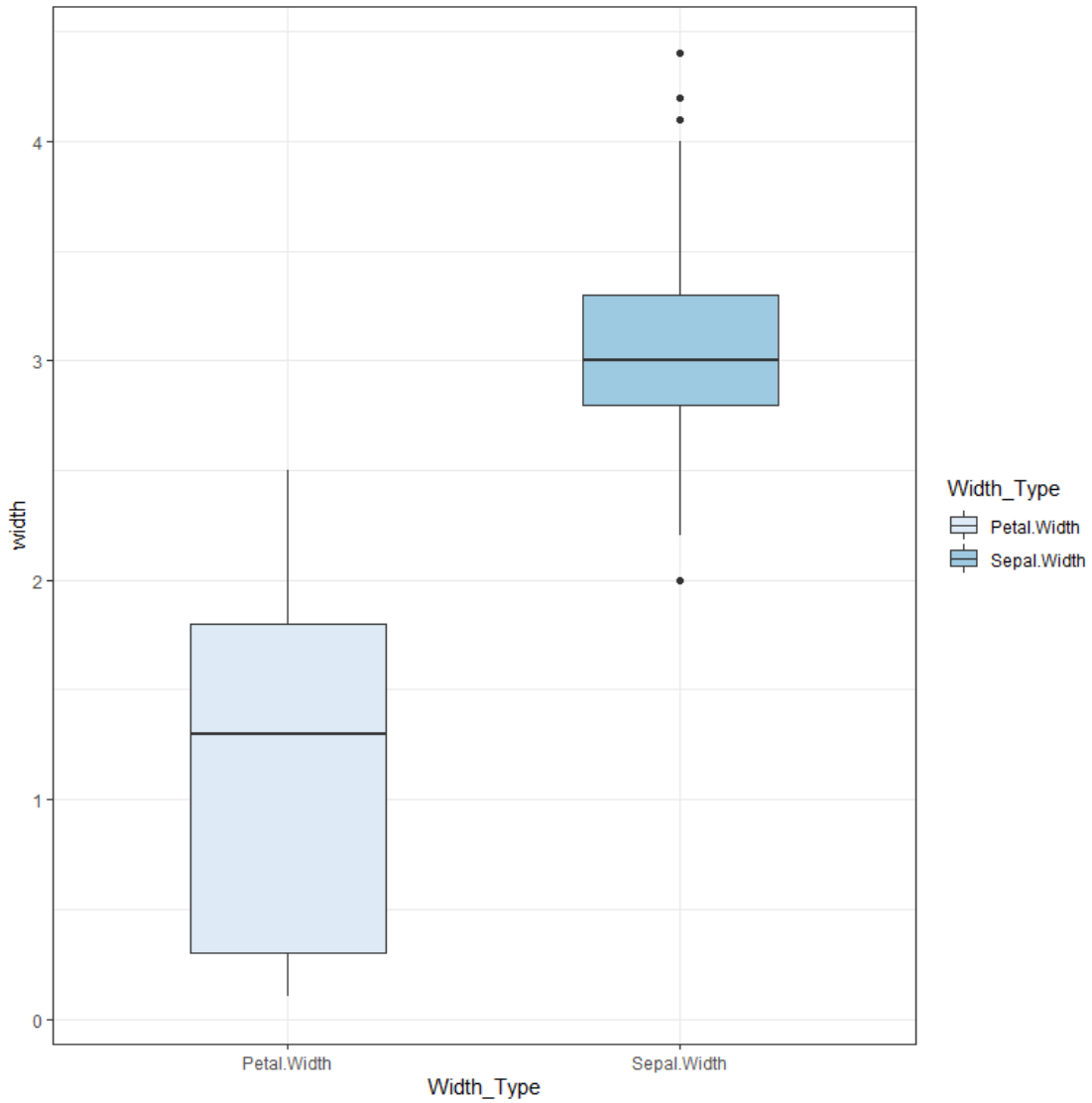
Construct new varibles:

```
iris_width<-iris %>% pivot_longer(c(Sepal.Width,Petal.Width),names_to = "Width_Type",values_to = "width")
head(iris_width)
```

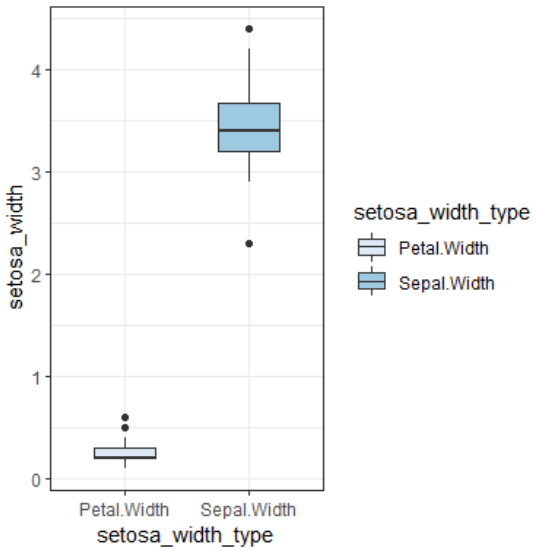
```
## # A tibble: 6 x 5
##   Sepal.Length Petal.Length Species Width_Type  width
##          <dbl>        <dbl> <fct>   <chr>       <dbl>
## 1          5.1          1.4 setosa  Sepal.Width   3.5
## 2          5.1          1.4 setosa  Petal.Width   0.2
## 3          4.9          1.4 setosa  Sepal.Width   3
## 4          4.9          1.4 setosa  Petal.Width   0.2
## 5          4.7          1.3 setosa  Sepal.Width   3.2
## 6          4.7          1.3 setosa  Petal.Width   0.2
```

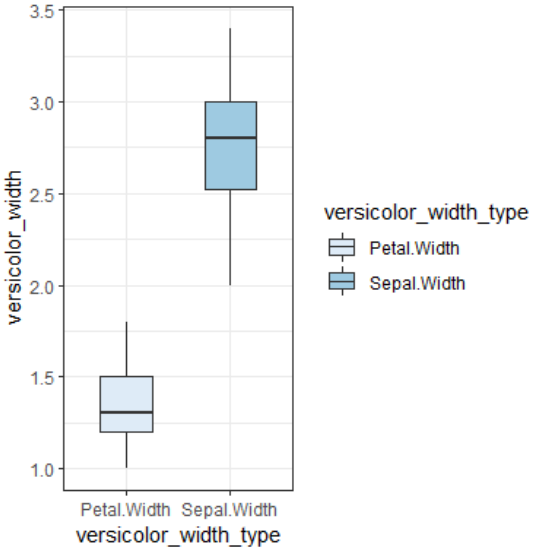
Conduct EDA:

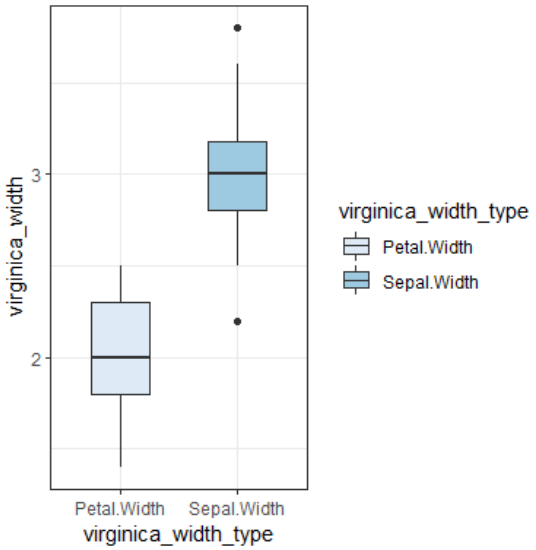
```
library(cowplot)
ggplot(iris_width,aes(Width_Type,width,fill=Width_Type))+geom_boxplot(width=0.5)+theme_bw()+scale_fill_brewer()
```




```
p1<-ggplot(iris_width[which(iris_width$Species %in% "setosa"),] %>% m
utate(setosa_width=width) %>%mutate(setosa_width_type=Width_Type),ae
s(setosa_width_type,setosa_width,fill=setosa_width_type))+geom_boxpl
ot(width=0.5)+theme_bw()+scale_fill_brewer()
p2<-ggplot(iris_width[which(iris_width$Species %in% "versicolor"),]
 %>% mutate(versicolor_width=width) %>%mutate(versicolor_width_type=
Width_Type),aes(versicolor_width_type,versicolor_width,fill=versicol
or_width_type))+geom_boxplot(width=0.5)+theme_bw()+scale_fill_brewer
()
p3<-ggplot(iris_width[which(iris_width$Species %in% "virginica"),]
 %>% mutate(virginica_width=width) %>%mutate(virginica_width_type=Wi
dth_Type),aes(virginica_width_type,virginica_width,fill=virginica_wi
dth_type))+geom_boxplot(width=0.5)+theme_bw()+scale_fill_brewer()
plist<-list()
plot_grid(p1,p2,p3)
```


setosa_width
4
3
2
1
0
Petal.Width
Sepal.Width
setosa_width_type
setosa_width_type
Petal.Width
Sepal.Width


versicolor_width
3.5
3.0
2.5
2.0
1.5
1.0
Petal.Width
Sepal.Width
versicolor_width_type
versicolor_width_type
Petal.Width
Sepal.Width


virginica_width
3
2
Petal.Width
Sepal.Width
virginica_width_type
virginica_width_type
Petal.Width
Sepal.Width

Homework5

gx

2022/8/10

R work

Use the dataset 'bridge.txt' to solve the following questions.

Question1:

Delete the variable 'Case' and transform all the variables to the log form (see the instructions provided in the Rmarkdown file.

R codes:

```
data<-read.table("D:\\bridge.txt",header=T)
data<-data[,-1]
data<-log(data,2)
head(data)
```

```
##       Time     DArea    CCost     Dwgs   Length Spans
## 1 6.300124 1.8479969 6.364572 2.584963 6.491853     0
## 2 8.273796 2.4141355 8.722124 3.584963 6.977280     1
## 3 7.527477 2.6530600 7.490249 3.169925 6.285402     0
## 4 6.121015 1.1375035 6.643856 2.321928 5.906891     0
## 5 6.104337 0.5260688 6.686501 2.321928 5.906891     0
## 6 6.580447 2.4329594 7.070389 2.321928 5.906891     0
```

Queston2:

Construct the full linear regression model using Time (in log form) as the response variable. Show the model summary.

R Codes:

```
fullmodel<-lm(Time ~., data)
summary(fullmodel)
```

```
##
## Call:
## lm(formula = Time ~ ., data = data)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -0.98671 -0.24767 -0.03757  0.33408  0.97104
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.29786    0.89340   3.691 0.000681 ***
## DArea       -0.04564    0.12675  -0.360 0.720705
## CCost        0.19609    0.14445   1.358 0.182426
## Dwgs         0.85879    0.22362   3.840 0.000440 ***
## Length      -0.03844    0.15487  -0.248 0.805296
## Spans        0.23119    0.14068   1.643 0.108349
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4529 on 39 degrees of freedom
## Multiple R-squared:  0.7762, Adjusted R-squared:  0.7475
## F-statistic: 27.05 on 5 and 39 DF,  p-value: 1.043e-11
```

Queston3:

Use stepwise selection with BIC to select the best model, show clearly what is 'k' used in the code, and show the model summary of the selected model.（Set trace = FALSE in the code so that we only see the last selected model.

R node:

```
library(tidyverse)
library(skimr)
library(MASS)
stepwise<-stepAIC(fullmodel,direction = "both",trace =F,k=log(nrow(d
ata)))
```

Summary:

```
summary(stepwise)

##
## Call:
## lm(formula = Time ~ Dwgs + Spans, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.99039 -0.35674 -0.08639  0.37582  0.91984
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  3.84007    0.38767   9.905 1.49e-12 ***
## Dwgs         1.04163    0.15420   6.755 3.26e-08 ***
## Spans        0.28530    0.09095   3.137  0.00312 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4479 on 42 degrees of freedom
## Multiple R-squared:  0.7642, Adjusted R-squared:  0.753
## F-statistic: 68.08 on 2 and 42 DF,  p-value: 6.632e-14
```

Queston4:

Give interpretations of the coefficients of the significant variables, remember we are using the log form of all the variables.

$$log_2(\hat{Y}) = 3.84 + 1.04 * log_2(Dwgs) + 0.29 * log_2(Spans)$$