Basic Template

```
setwd("/Users/kurisuuu/Downloads")
focus=read.csv("focus.csv")
set.seed(1)
inds <- sample(1:nrow(focus),500) #indices

ford.focus <- focus[inds,] %>%
  as_tibble() %>%
  select(-model) %>%
  mutate_if(is.character,as.factor) %>%
  mutate(fuelType=as_factor(fuelType),engineSize=as.numeric(engineSize)) %>%
  select(price,everything())

options(knitr.duplicate.label = "allow")
```

Introduction

Exploratory Data Analysis

```
ford.focus %>%
  summarise(n=n(),Mean=round(mean(price),digits=1), St.Dev=round(sd(price),digits=1),
    Min=min(price), Q1 = quantile(price,0.25), Median=median(price),
    Q3 = quantile(price,0.75), Max=max(price)) %>%
kable(caption = '\\label{tab:price} Summary statistics on
  price of 500 UK Used Car Data set.') %>%
kable_styling(latex_options = "hold_position")
```

Table 1: Summary statistics on price of 500 UK Used Car Data set.

n	Mean	St.Dev	Min	Q1	Median	Q3	Max
500	14343.6	4521.6	1850	11337.5	14496.5	17491.25	28930

Table 1 shows that the summaries of price of 500 UK Used Car Data set. For example the mean price 13480.5 pounds. We also note that the variability in the price as 4784.7 pounds. The messages can be easily seen the in the following boxplot which summarise the distribution of car price.

```
ggplot(ford.focus, aes(y = price)) +
  geom_boxplot() +
  labs(x = "Ford", y = "Price",
    title = "price of 500 UK Used Car Data set")
```

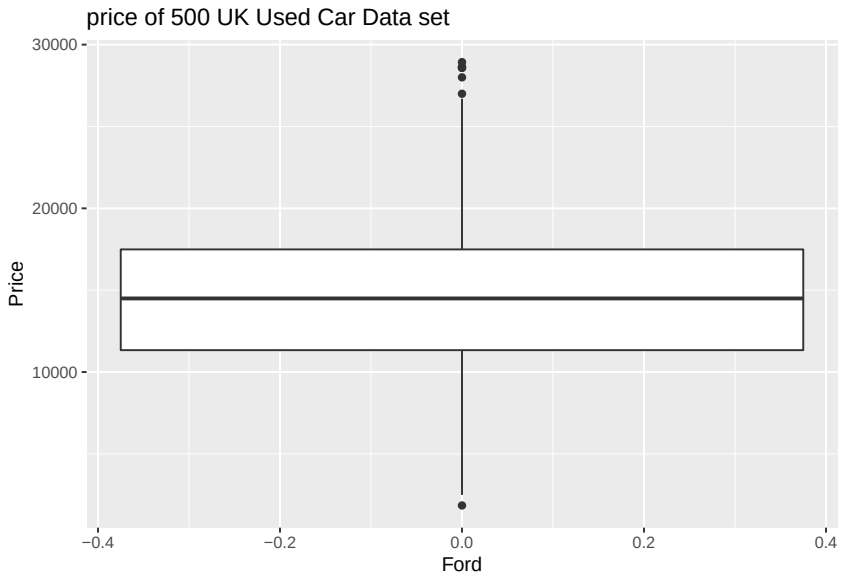

Figure 1: Price of 500 used Ford cars.

```
ford.focus %>%
  summarise(n=n(),Mean=round(mean(mileage),digits=1), St.Dev=round(sd(mileage),digits=1),
    Min=min(mileage), Q1 = quantile(mileage,0.25), Median=median(mileage),
    Q3 = quantile(mileage,0.75), Max=max(mileage)) %>%
kable(caption = '\\label{tab:milage} Summary statistics on
  mileage of 500 UK Used Car Data set.') %>%
kable_styling(latex_options = "hold_position")
```

Table 2: Summary statistics on mileage of 500 UK Used Car Data set.

n	Mean	St.Dev	Min	Q1	Median	Q3	Max
500	19477.5	17406	1	8228.75	13990.5	24434.5	93400

```
ford.focus %>%
  summarise(n=n(),Mean=round(mean(engineSize),digits=1), St.Dev=round(sd(engineSize),digits=1),
    Min=min(engineSize), Q1 = quantile(engineSize,0.25), Median=median(engineSize),
    Q3 = quantile(engineSize,0.75), Max=max(engineSize)) %>%
kable(caption = '\\label{tab:engineSize} Summary statistics on
  engineSize of 500 UK Used Car Data set.') %>%
kable_styling(latex_options = "hold_position")
```

Table 3: Summary statistics on engineSize of 500 UK Used Car Data set.

n	Mean	St.Dev	Min	Q1	Median	Q3	Max
500	1.4	0.4	0	1	1.2	1.6	2.3

Formal Data Analysis

Conclusions