```r
#a) Load the data set:

house_data <- tibble(
  Area = c(800, 1000, 1200, 1500, 1800, 2000, 2200, 2500, 2800, 3000),
  Price = c(25, 30, 36, 45, 52, 58, 65, 72, 80, 88)
)


print(house_data)
```

```
A tibble: 10 × 2
   Area Price
  <dbl> <dbl>
1   800    25
2  1000    30
3  1200    36
4  1500    45
5  1800    52
6  2000    58
7  2200    65
8  2500    72
9  2800    80
0  3000    88
```

```r
#b) Description of our data

house_data %>%
  summarise(
    Count = n(),
    Mean_Area = mean(Area),
    SD_Area = sd(Area),
    Min_Area = min(Area),
    Max_Area = max(Area),
    Mean_Price = mean(Price),
    SD_Price = sd(Price),
    Min_Price = min(Price),
    Max_Price = max(Price)
  )
```

```
A tibble: 1 × 9
Count Mean_Area SD_Area Min_Area Max_Area Mean_Price SD_Price Min_Price Max_Price
<int>     <dbl>   <dbl>    <dbl>    <dbl>      <dbl>    <dbl>     <dbl>     <dbl>
   10      1880    757.      800     3000       55.1     21.4        25        88
```

```r
#c) Explore data

ggplot(house_data, aes(x = Area, y = Price)) +
  geom_point(color = "blue", size = 3) +
  labs(title = "House Price vs Area",
       x = "Area (sqft)",
       y = "Price (lakh Rupees)") +
  theme_minimal()

#d) Define a linear model

model <- lm(Price ~ Area, data = house_data)


#e) Plot the regression line

ggplot(house_data, aes(x = Area, y = Price)) +
  geom_point(color = "blue", size = 3) +
  geom_smooth(method = "lm", se = FALSE, color = "red", linewidth = 1.2) +
  labs(title = "Linear Regression: Price ~ Area",
       x = "Area (sqft)",
       y = "Price (lakh Rupees)") +
  theme_minimal()
```

```
geom_smooth()` using formula = 'y ~ x'
```

```r
#f) summary of the model

summary(model)
```

```
lm(formula = Price ~ Area, data = house_data)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0477 -0.5610 -0.1008  0.5532  1.3115

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.0764158  0.7328147   2.833    0.022 *
Area        0.0282040  0.0003641  77.454 8.61e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8268 on 8 degrees of freedom
Multiple R-squared:  0.9987,    Adjusted R-squared:  0.9985
F-statistic:  5999 on 1 and 8 DF,  p-value: 8.606e-13

>
> #g) The regression was run on 10 observations
>
> #h) What is the R-squared
>
> print(paste("R squared error = ", summary(model)$r.squared))
[1] "R squared error =  0.998668251944291"
>
> #i) Is Size statistically significant predictor?
>
> summary(model)$coefficients
              Estimate   Std. Error  t value     Pr(>|t|)
(Intercept) 2.07641583 0.7328146849  2.83348 2.203189e-02
Area        0.02820403 0.0003641385 77.45413 8.605534e-13
>
> #j) Regression equation
>
> #Price = 2.123 + 0.0285 × Area
> #Price = coeff. * area + intercept
>
```