Problem One - Statement of Problem

This question, using the provided measured data set, which contains the eruption time spans and wait time for the subsequent eruption.Using the relationship between these two data to observe the number of Old Faithful eruptions in Yellowstone. Create a graph through R language.

Methods

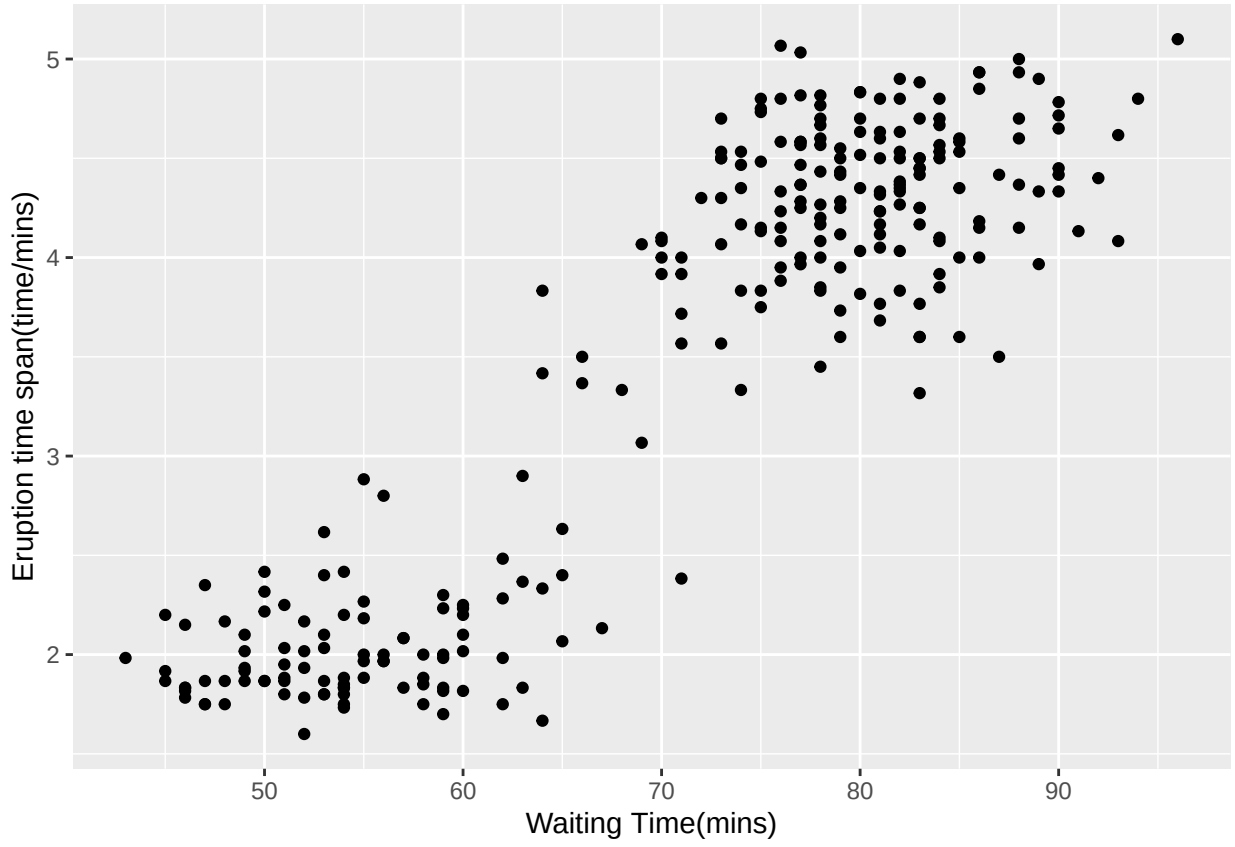
Load the faithful data as a tibble, Using ggplot to create the graph framework, in this observation, I chose to use subsequent wait time as the X axis of my chart, and the Y axis is the error time span. I think this design can help me better see what the number of fathful eruptions at each interval is. So by mapping = aes() to determine the data that my XY axis will take. Use geom_point() to represent each eruption as a plot.

Create a linear trend line to the plots by using the function of geom_smooth() and use the method = lm.

Results

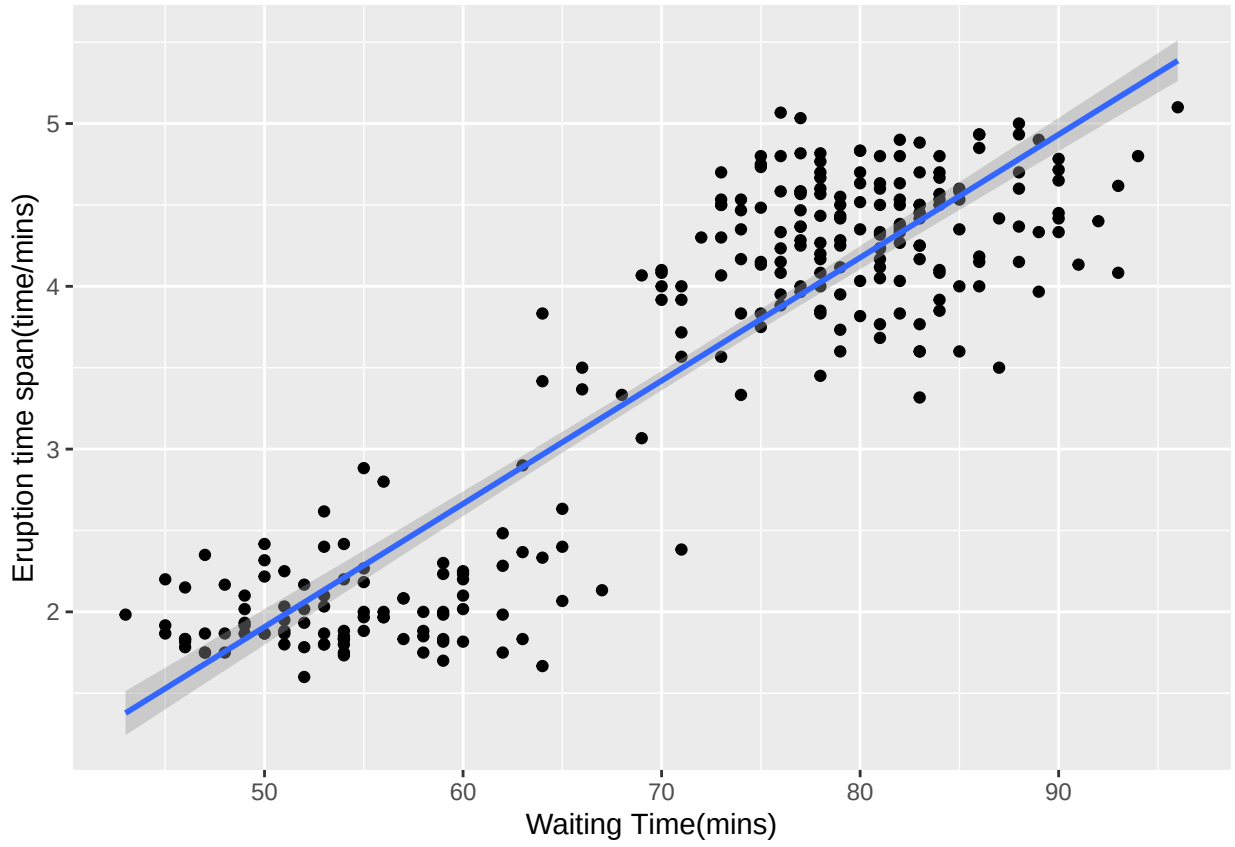
```
faithful <- as_tibble(faithful)
```

```
ggplot(data = faithful) +
  geom_point(mapping = aes(x = waiting, y = eruptions)) +
  xlab("Waiting Time(mins)") + ylab("Eruption time span(time/mins)")
```



```
ggplot(data = faithful, mapping = aes(x = waiting, y = eruptions)) +
  geom_point() +
  geom_smooth(method = lm) +
  xlab("Waiting Time(mins)") + ylab("Eruption time span(time/mins)")
```

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



Conclusion(s)

It can be seen from the graph that when the waiting time of faithful is longer, the number of eruptions time will gradually increase.

Appendix - R code

```
#knitr::opts_chunk$set(echo = TRUE)
#library(tidyverse)
```

(1). ggplot(data = faithful) + geom_point(mapping = aes(x = waiting, y = eruptions)) + xlab("Waiting Time(mins)") + ylab("Eruption time span(time/mins)")

(2). ggplot(data = faithful, mapping = aes(x = waiting, y = eruptions)) + geom_point() + geom_smooth(method = lm) + xlab("Waiting Time(mins)") + ylab("Eruption time span(time/mins)")

Problem Two - Statement of Problem

Use the species characteristics provided by the built-in iris dataset to determine different species of iris flowers, including the length of the sepals, the width of the sepals, the length of the petals, and the width of the petals. The three different iris varieties are Setosa, Versicolor and Virginica. Build the graph using the R language and choose two parameters that we think are useful to analyze.

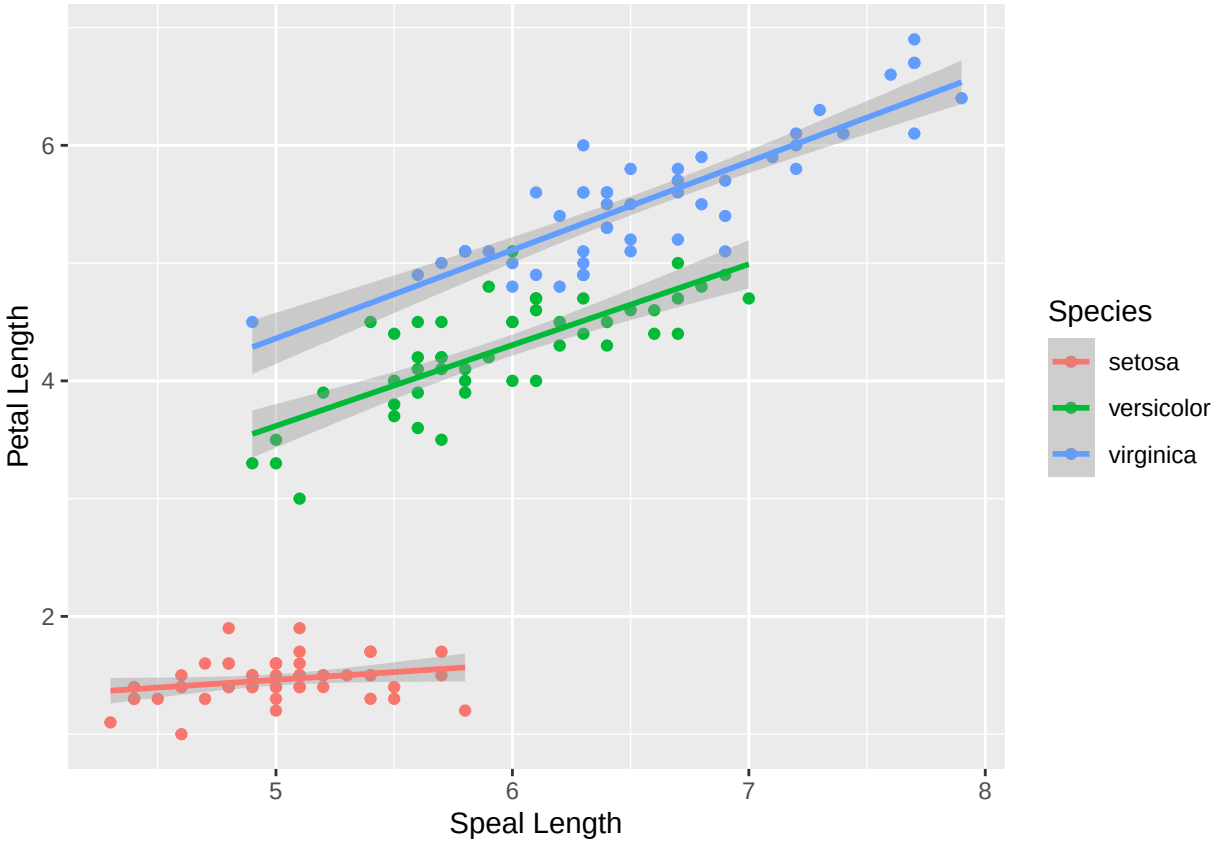
Method

On this issue, I first looked at the overall data and found that there was a clear difference between the parameters of the length of sepal and the length of petal in the Iris species, so I chose these two parameters as my XY. First, using ggplot(), select the dataset (data = iris) of iris, and decide the parameters of the XY axis (x = Sepal.Length, y = Petal.Length), and then use Species to distinguish between different varieties of Iris (color= Species). Each corresponding symbol is represented by a plot (geom_point()), and then the linear of these plots is drawn with geom_smooth(), but what we need is the linear trend line, so we equal the method to lm in geom_smooth().

Results

```
ggplot(data = iris, mapping = aes(x = Sepal.Length, y = Petal.Length, color = Species)) +
  geom_point() +
  geom_smooth(method = lm) +
  xlab("Speal Length") + ylab("Petal Length")
```

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



Conclusion(s)

Yes, there are significant differences between Setosa and the other two species of Iris. There are obvious differences between the iris of Setosa and the other two varieties. in the graph, I define the length of the sepal of Iris on the X-axis, and the length of petal of Iris on the Y-axis. Through the graph, we can see the length of the sepal and petal of the species Setosa, both are short than the other. However, the lengths of Versicolor and Virginica are long, but after the speal length is large than 7, only Virginica this species.

Appendix - R code

(1). ggplot(data = iris, mapping = aes(x = Sepal.Length, y = Petal.Length, color = Species)) + geom_point() + geom_smooth(method = lm) + xlab("Speal Length") + ylab("Petal Length")