Adv_R_Homework1

Zixuan Gao 23206703

1. Introduction

The **'Orange'** is a dataset which is available in R. This dataset has 35 rows and 3 columns of records of the growth of orange trees. This dataset was originally part of package **nlme**, and that has methods *(including for [, as.data.frame, plot and print)* for its grouped-data classes.It is extracted from Draper, N. R. and Smith, H. (1998), *Applied Regression Analysis (3rd ed)*, Wiley (exercise 24.N) & Pinheiro, J. C. and Bates, D. M. (2000) *Mixed-effects Models in S and S-PLUS*, Springer.(See Orange for more details).

This analysis explores the **relationship between orange tree age and circumference.**

Data loading

Below the first 6 rows of the dataset are displayed:

```
  Tree  age circumference
1    1  118            30
2    1  484            58
3    1  664            87
4    1 1004           115
5    1 1231           120
6    1 1372           142
```

Description

Below are some descriptive statistics of the tree age and circumference in **Orange** dataset.

```
Tree       age         circumference
3:7   Min.   : 118.0   Min.   : 30.0
1:7   1st Qu.: 484.0   1st Qu.: 65.5
5:7   Median :1004.0   Median :115.0
2:7   Mean   : 922.1   Mean   :115.9
4:7   3rd Qu.:1372.0   3rd Qu.:161.5
      Max.   :1582.0   Max.   :214.0
```

2. Data visualisation

We can plot the data and visually analyze it using the R package *ggplot2*:

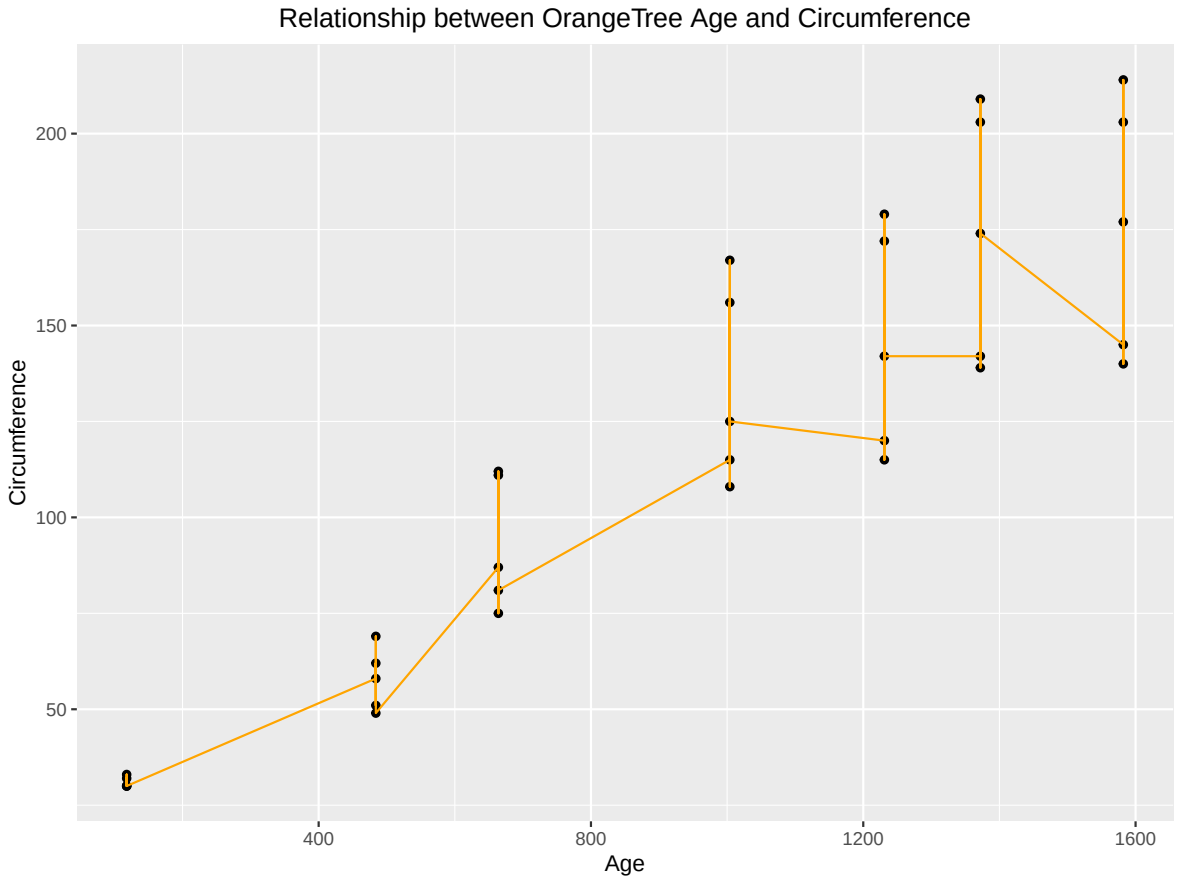

Figure 1: Relationship between OrangeTree Age and Circumference in Orange dataset

3. Summary

From the line graph, we can see that the age of the orange tree is positively correlated with its circumference. The older the tree is, the larger its circumference is. However, as the tree ages, the differences in circumference also increase. Trees of the same age may have very different circumferences. For example, when the tree is close to 1600 years old, its circumference is less than 150 at the minimum and greater than 220 at the maximum.