

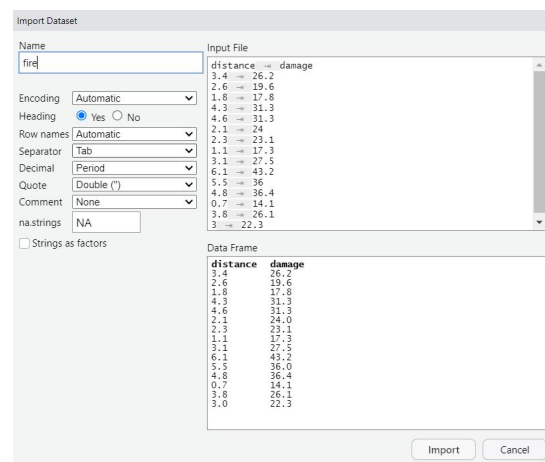
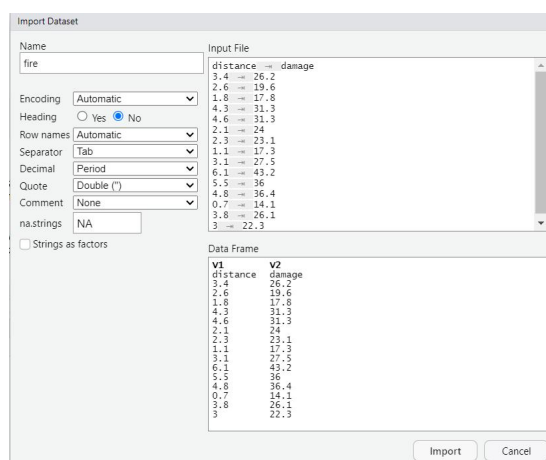
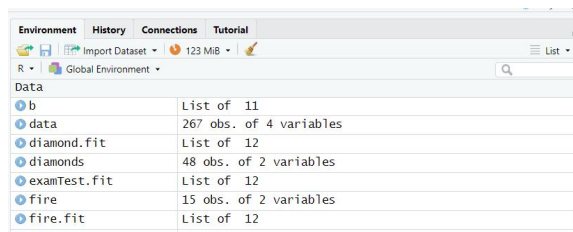
一 导入数据：

将数据导入 R 成为一个数据框架。

书上使用的方法为采用 `read.table()` 函数。

```
> ## Importing data into R
> Stats20x.df = read.table("Data/STATS20x.txt", header=TRUE)
```

我推荐采取 GUI 导入数据的方法，从 Environment 里的 Import Dataset 直接导入数据



注意 Heading 选 Yes 和 No 的区别

导入数据后可以用 `dim`（数据框名）来查看数据的行列维度

注意导入数据后选择有助于记忆的名字，例如可以使用 `.df` 后缀（数据框架）

注意 R 语言对大小写敏感

二．数据框架

R 语言数据查询：[数据框架.列\[索引\]](#)

查看某一数值列的分布采用直方图 `hist（数据框架.列，xlabel=，ylabel=）`

三．Summary 函数

Summary 自动分析函数：

对于数值型列：直接计算出数据框架的最小值、四分位数、中位数、均值、最大值

对于因子型（分类变量）列：显示每个分类的频数

对于字符型列：显示字符的长度统计

四．简单直线拟合

绘制散点图：`plot(列 y~列 x，data=数据框架)`

绘制散点图上加入平滑趋势线：`trendscatter（列 y~列 x，data=数据框架）`

从图中直接看回归的趋势

如果是简单的线性回归，可以使用 `lm()` 函数来估计线性回归参数：`a.fit<-lm(列 y~列 x，data=数据框架)`，`Summary（a.fit）` 查看模型详细拟合结果

Coef () 函数可以提取模型拟合参数: **Coef (a.fit)**

最后进行预测:

Predic<-predict(a.fit,newdata=data.frame(列 x=c(point1,point2,...,pointn)))、

五. 总结第一章常用代码:

Most of the R-code you need for this chapter

Creating a data frame by importing the text file Data/STATS20x.txt from the subfolder Data:

```
> Stats20x.df = read.table("Data/STATS20x.txt", header=TRUE)
```

Plotting a scatter plot of y (Exam) vs. x (Test):

```
> plot(Exam~Test, data=Stats20x.df)
```

and/or a trend-scatter-plot:

```
> trendscatter(Exam ~Test, data = Stats20x.df)
```

Fitting a linear (straight line) model and getting the estimated values:

```
> examtest.fit=lm(Exam~Test, data=Stats20x.df)
> summary(examtest.fit)
```

Adding your fitted/predicted blue-dashed line to the scatter-plot created above:

```
> abline(examtest.fit, lty = 2, col = "blue")
```

还有预测函数 predict ()