Goals: 1. To begin formulating the types of questions one might ask concerning the relationship between two variables. As well, to be able to identify, for a given question, which variable is being used in an explanatory role, and which is seen as the response.

2. To be able to produce plots and/or tables that are useful in exploring whether there is a relationship between two variables (either 2 categorical, or 1 categorical and 1 quantitative).

3. To be able to glean from those plots/tables a preliminary answer to the “Is there a relationship?” question.

Read: Section 4.3 of LAS

Useful R commands:

<table>
<thead>
<tr>
<th>Commands</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>quantile()</td>
<td>Find the position of a quantile in a dataset</td>
</tr>
<tr>
<td>summary()</td>
<td>When Hmisc package loaded, produces tables like on pp. 416–418</td>
</tr>
<tr>
<td>xtabs()</td>
<td>Cross-tabulation tables using formulas</td>
</tr>
<tr>
<td>boxplot()</td>
<td>R (built-in) command for producing boxplots</td>
</tr>
<tr>
<td>bwplot()</td>
<td>Lattice command for producing boxplots</td>
</tr>
</tbody>
</table>

Examples (without output):

```r
> ss = read.table('http://www.calvin.edu/~scofield/data/tab/ssurv.txt',
  sep='\t', header=T)
> median(ss$cds, na.rm = T)
> mean(ss$cds, na.rm = T)
> var(ss$cds, na.rm = T)
> sd(ss$cds, na.rm = T)
> summary(ss$cds)

> table(ss$smoker, ss$region)
> xtabs(~ smoker + region, data = ss)

> boxplot(ss$hourssleep)
> boxplot(hourssleep ~ gender, data=ss)
> library(lattice)
```
> bwplot(~ hourssleep | gender, data=ss)
> bwplot(hourssleep ~ gender, data=ss)
> histogram(~ hourssleep | gender, data=ss)

> summary(hourssleep ~ gender, data=ss)
> library(Hmisc)
> summary(hourssleep ~ gender, data=ss)
> summary(hourssleep ~ gender, data=ss, fun=mean)
> summary(hourssleep ~ gender, data=ss, fun=median)

Terms to know: Simpson’s paradox