Topic 5.1 Overview of ANOVA

Loaded needed packages.

```
library(Stat2Data)
library(mosaic)
```

EXAMPLE 5.1 Fat rats: A randomized experiment

Create a dataframe for **FatRats** and look at the structure of the data.

```
data("FatRats")
str(FatRats)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ Gain   : int  73 102 118 104 81 107 100 87 117 111 ...
##  $ Protein: Factor w/ 2 levels "Hi","Lo": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Source : Factor w/ 3 levels "Beef","Cereal",..: 1 1 1 1 1 1 1 1 1 1 ...
```

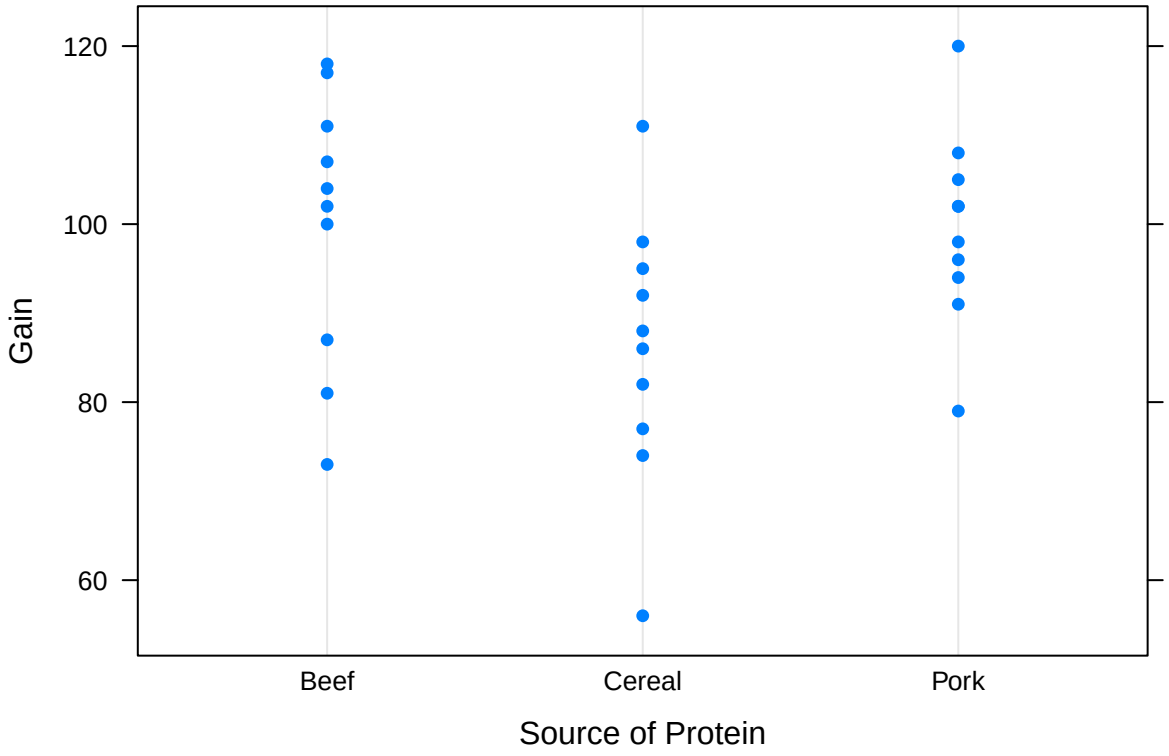
Create a subset with just the 30 rats having Protein="Hi."

```
FatRatsHi=subset(FatRats,Protein=="Hi")
```

FIGURE 5.1 Weight gain versus protein source for baby rats on a high protein diet

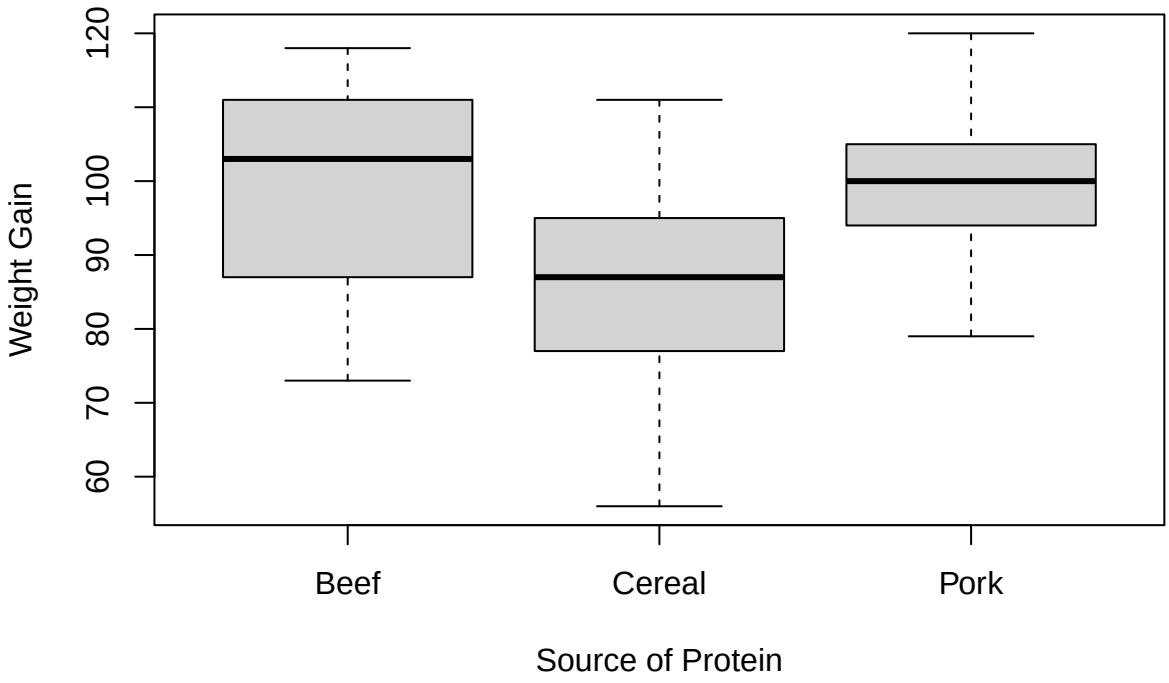
(a) Dotplots

```
dotplot(Gain~Source,data=FatRatsHi,xlab="Source of Protein")
```



(b) Boxplots

```
boxplot(Gain~Source,data=FatRatsHi,xlab="Source of Protein",ylab="Weight Gain")
```



Find the mean and standard deviation for the overall sample.

```
favstats(~Gain,data=FatRatsHi)
```

```
##  min    Q1 median     Q3 max     mean      sd  n missing
##   56 86.25     97 104.75 120 95.13333 14.9083 30       0
```

Find the mean and standard deviation within each Source group.

```
favstats(Gain~Source,data=FatRatsHi)
```

```
##   Source min    Q1 median     Q3 max  mean       sd  n missing
## 1   Beef  73 90.25    103 110.00 118 100.0 15.13642 10       0
## 2 Cereal  56 78.25     87  94.25 111  85.9 15.02184 10       0
## 3   Pork  79 94.50    100 104.25 120  99.5 10.91635 10       0
```

EXAMPLE 5.2 Leafhopper diets

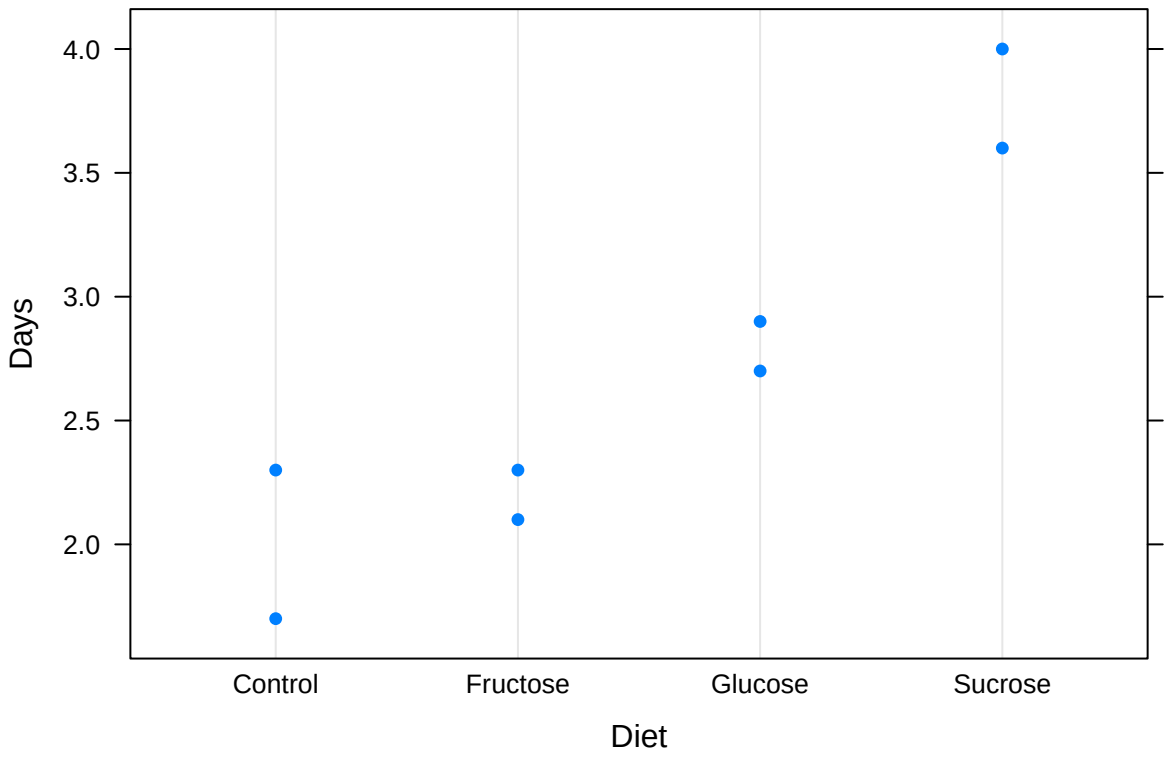
Create a dataframe for **Leafhoppers** and look at the structure of the data.

```
data("Leafhoppers")
str(Leafhoppers)
```

```
## 'data.frame':    8 obs. of  3 variables:
##  $ Dish: int  1 2 3 4 5 6 7 8
##  $ Diet: Factor w/ 4 levels "Control","Fructose",..: 1 1 4 4 3 3 2 2
##  $ Days: num  2.3 1.7 3.6 4 2.9 2.7 2.1 2.3
```

FIGURE 5.2 Survival time in days versus diet

```
dotplot(Days~Diet,data=Leafhoppers,xlab="Diet")
```



EXAMPLE 5.3 Teen pregnancy and the Civil War

Create a dataframe for **TeenPregnancy** and look at the structure of the data.

Note: The factor( ) in the second line reorders the state categories to match the order in the text. Otherwise R defaults to alphebetical order for the groups.

```
data("TeenPregnancy")
TeenPregnancy$CivilWar=factor(TeenPregnancy$CivilWar,levels=c("C","B","U","O"))
str(TeenPregnancy)
```

```
## 'data.frame':    50 obs. of  4 variables:
##  $ State   : Factor w/ 50 levels "AK","AL","AR",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ CivilWar: Factor w/ 4 levels "C","B","U","O": 4 1 1 4 3 4 3 3 1 1 ...
##  $ Church  : int  26 46 45 33 28 25 25 35 32 39 ...
##  $ Teen    : int  64 62 73 60 59 50 44 67 60 64 ...
```

TABLE 5.1 Mean teen pregnancy rate by Civil War status of states

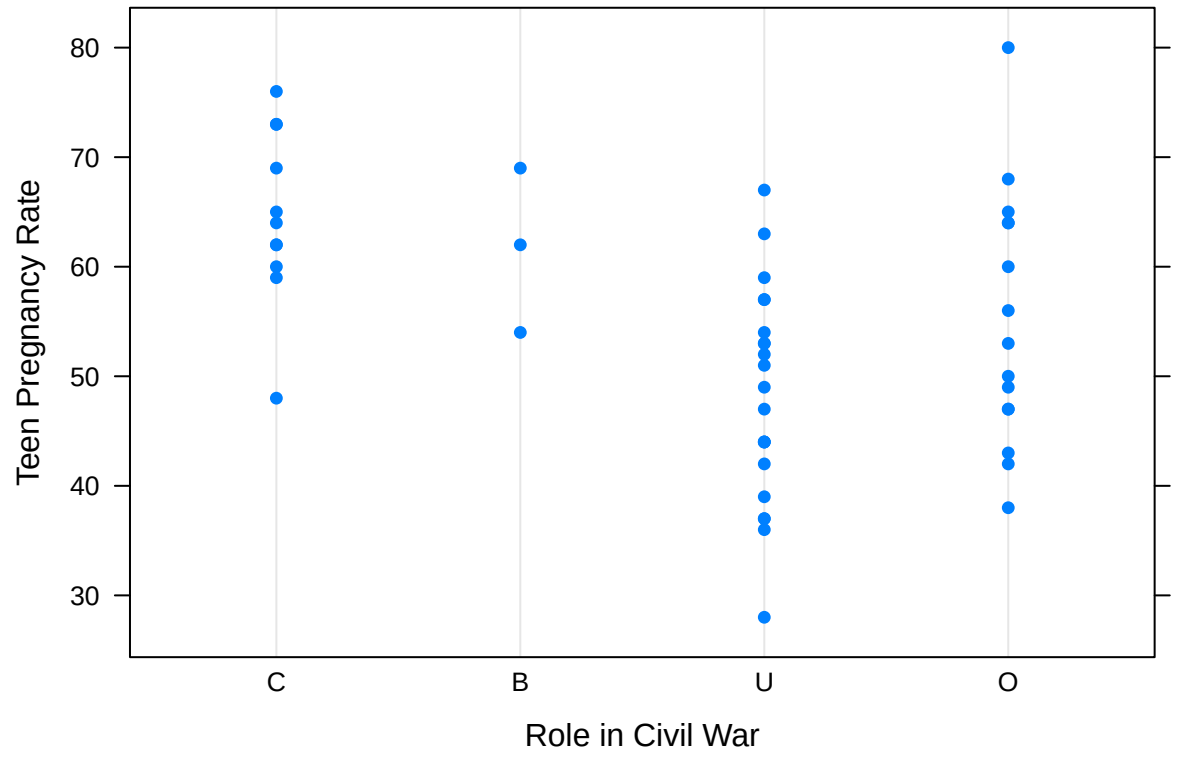
```
favstats(Teen~CivilWar,data=TeenPregnancy)
```

```
##   CivilWar min Q1 median   Q3 max     mean        sd  n missing
## 1        C  48 61     64 71.0  76 64.63636  7.953273 11       0
## 2        B  54 58     62 65.5  69 61.66667  7.505553  3       0
## 3        U  28 42     49 54.0  67 48.23810  9.828045 21       0
## 4        O  38 47     53 64.0  80 55.06667 11.578716 15       0
```

FIGURE 5.3 Plots of teen pregnancy rate versus role in the U.S. Civil War

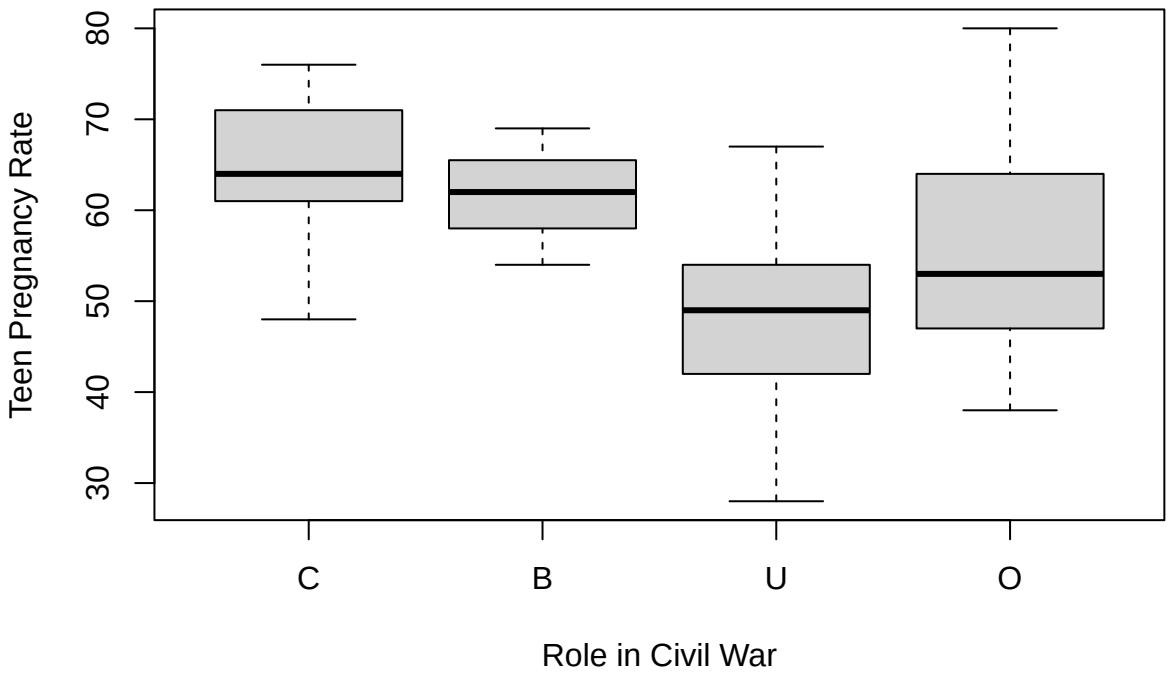
(a) Dotplots

```
dotplot(Teen~CivilWar,data=TeenPregnancy,xlab="Role in Civil War", ylab="Teen Pregnancy Rate")
```



(b) Boxplots

```
boxplot(Teen~CivilWar,data=TeenPregnancy,xlab="Role in Civil War", ylab="Teen Pregnancy Rate")
```


Teen Pregnancy Rate
30
40
50
60
70
80
C
B
U
O
Role in Civil War