Section 5.4 Formal Inference: Assessing and Using the ANOVA Model

Loaded needed packages.

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

EXAMPLE 5.10 Leafhoppers and psychotherapy: P-values

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
```

Get Leafhopper ANOVA table (with P-value).

```
AnovaModel=aov(Days~Diet,data=Leafhoppers)
summary(AnovaModel)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## Diet         3   3.92   1.307   17.42 0.00925 **
## Residuals    4   0.30   0.075
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

EXAMPLE 5.11 Fruit fly lifetimes

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

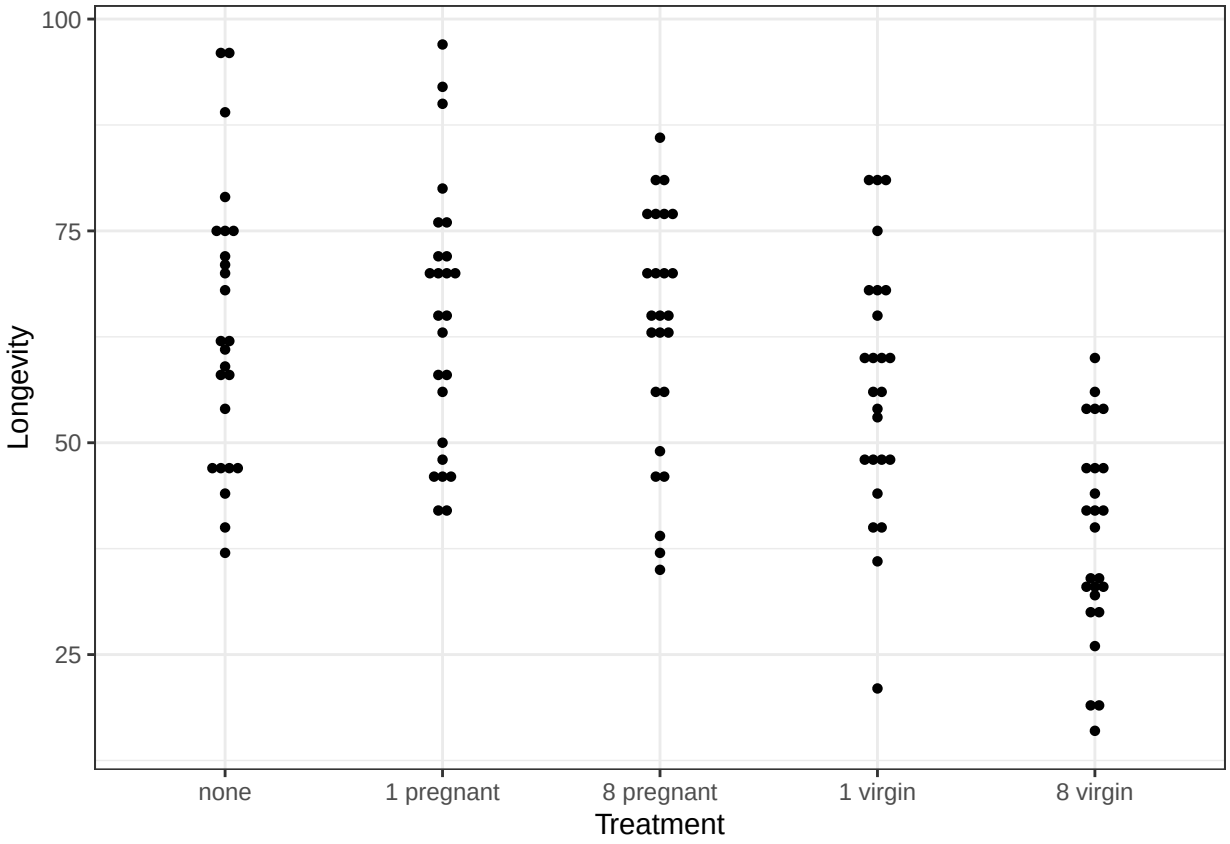
```
data("FruitFlies")
#order the Treatment categories
FruitFlies$Treatment=factor(FruitFlies$Treatment,levels=c("none","1 pregnant","8 pregnant","1 virgin","8
str(FruitFlies)
```

```
## 'data.frame':    125 obs. of  7 variables:
##  $ ID       : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Partners : int  8 8 8 8 8 8 8 8 8 8 ...
##  $ Type     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Longevity: int  35 37 49 46 63 39 46 56 63 65 ...
##  $ Thorax   : num  0.64 0.68 0.68 0.72 0.72 0.76 0.76 0.76 0.76 0.76 ...
##  $ Sleep    : int  22 9 49 1 23 83 23 15 9 81 ...
##  $ Treatment: Factor w/ 5 levels "none","1 pregnant",..: 3 3 3 3 3 3 3 3 3 3 ...
```

FIGURE 5.13 Fruit fly longevity by treatment groups

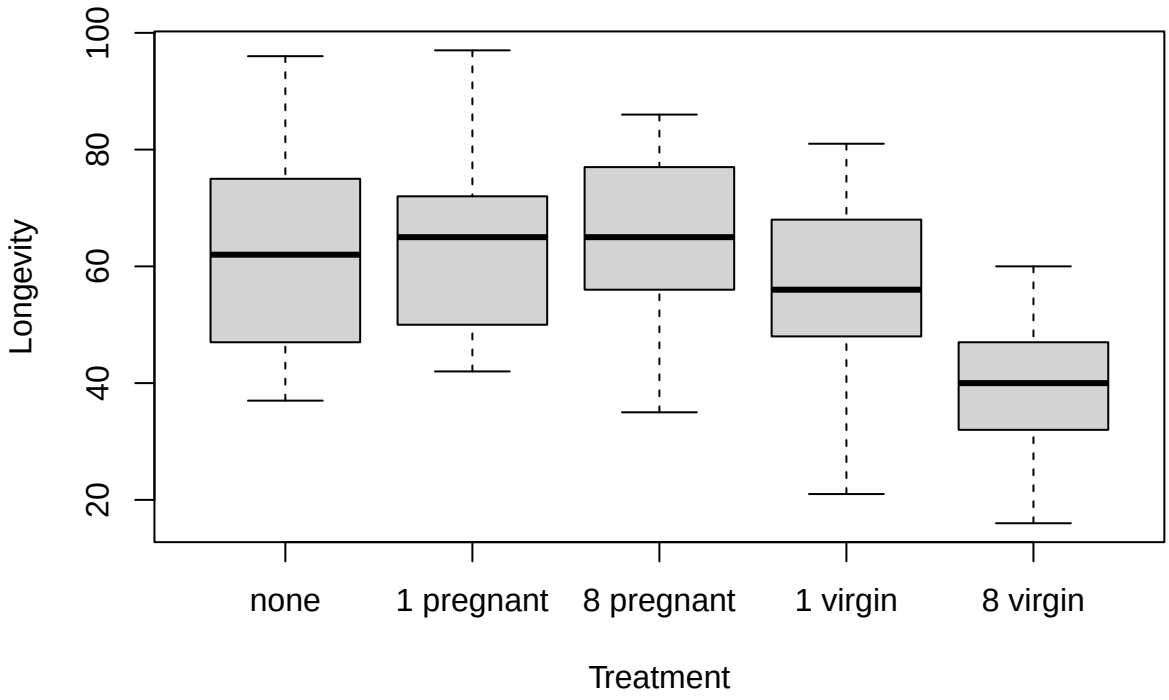
(a) Dotplots

```
ggplot(FruitFlies, aes(x = Treatment, y = Longevity)) +
  geom_dotplot(binaxis = "y", stackdir = "center",binwidth=1) +
  labs(x="Treatment") +theme_bw()
```



(a) Boxplots

```
boxplot(Longevity~Treatment,data=FruitFlies,ylab="Longevity",xlab="Treatment")
```



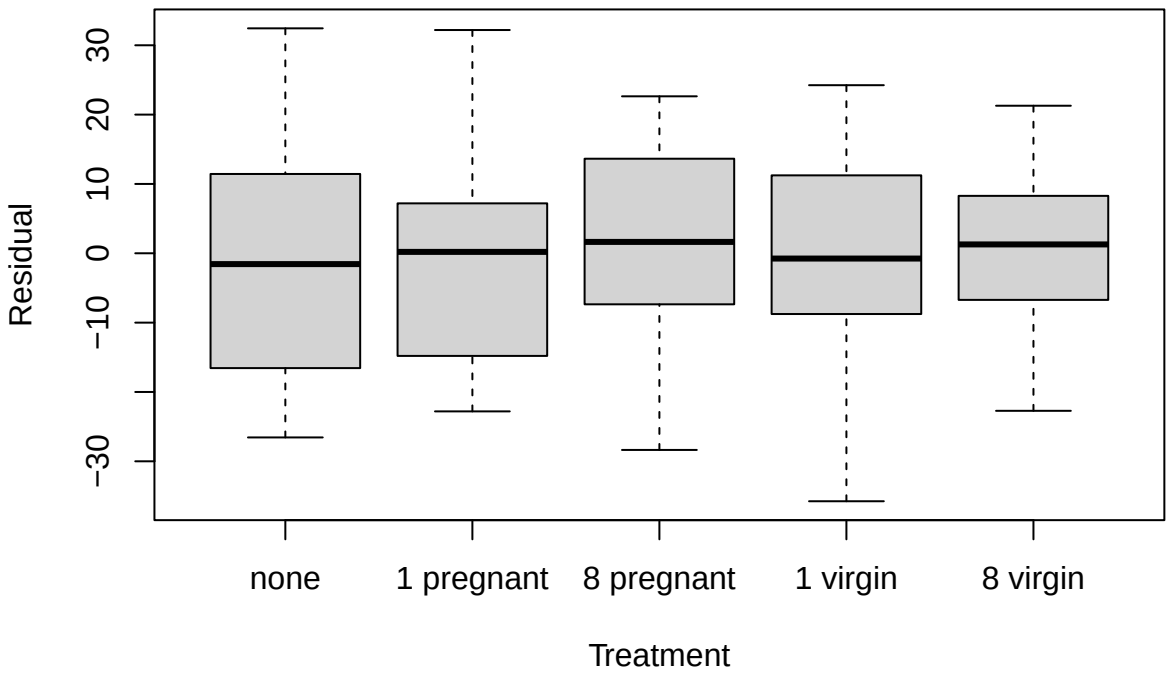
Fit the one-way ANOVA model.

```
FFmodel=aov(Longevity~Treatment,data=FruitFlies)
```

FIGURE 5.14 Residual plots for fruit fly lifetime ANOVA

(a) Boxplots of residuals

```
boxplot(FFmodel$residuals~Treatment,data=FruitFlies,ylab="Residual",xlab="Treatment")
```



(b) Normal quantile plot of residuals

```
qqnorm(FFmodel$residuals,ylab="Residuals",xlab="Normal Quantiles",main="")
qqline(FFmodel$residuals)
```

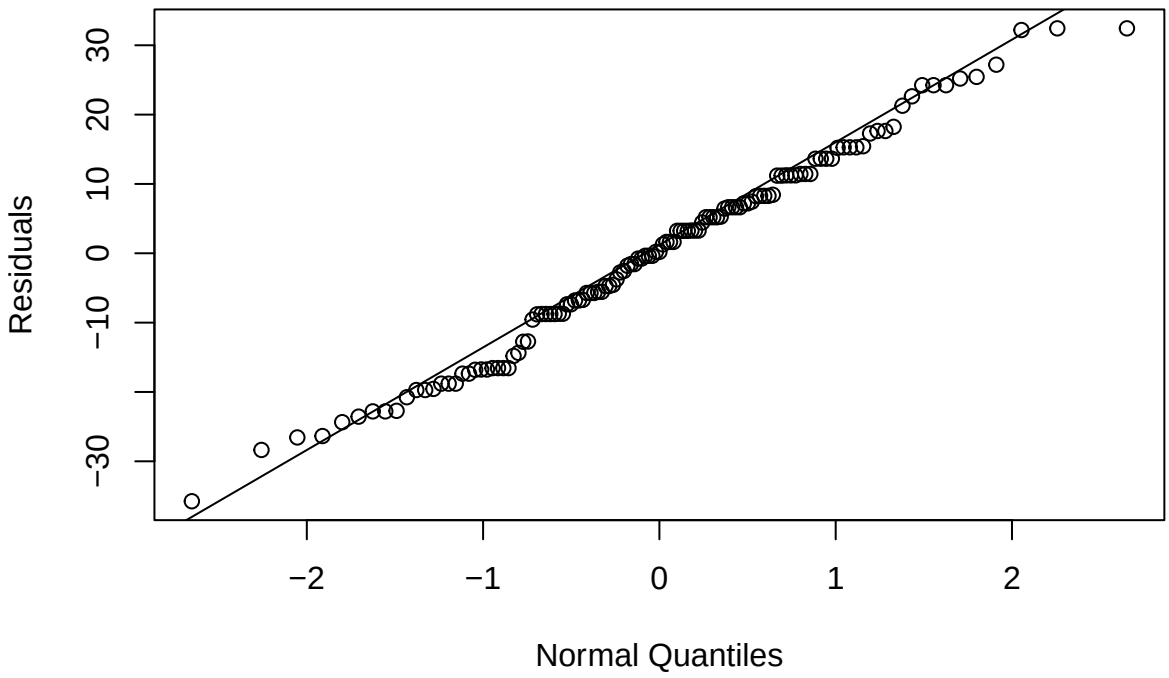


TABLE 5.2 Fruit fly lifetime statistics; overall and broken down by groups

Compute mean and standard deviation, by Groups.

```
favstats(Longevity~Treatment,data=FruitFlies)[c("Treatment","n","mean","sd")]
```

```
##    Treatment  n  mean       sd
## 1       none 25 63.56 16.45215
## 2 1 pregnant 25 64.80 15.65248
## 3 8 pregnant 25 63.36 14.53983
## 4   1 virgin 25 56.76 14.92838
## 5   8 virgin 25 38.72 12.10207
```

Find the overall statistics.

```
favstats(~Longevity,data=FruitFlies)[c("n","mean","sd")]
```

```
##    n  mean       sd
##  125 57.44 17.56389
```

Estimate the treatment effects for FruitFlies.

```
mean(Longevity~Treatment,data=FruitFlies)-mean(FruitFlies$Longevity)
```

```
##       none 1 pregnant 8 pregnant   1 virgin   8 virgin
##       6.12       7.36       5.92      -0.68     -18.72
```

Get the ANOVA table for FruitFlies.

```
summary(FFmodel)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Treatment     4  11939  2984.8   13.61 3.52e-09 ***
## Residuals   120  26314   219.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

EXAMPLE 5.13 Church attendance and teen pregnancy

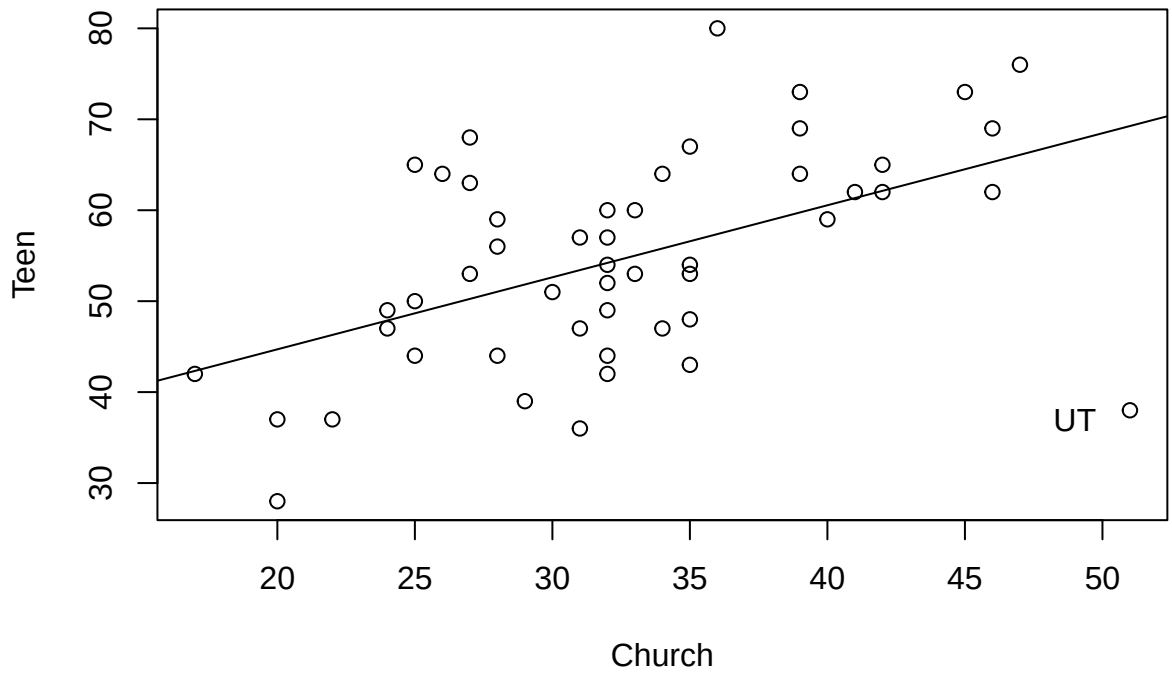
Create a dataframe for **TeenPregnancy** and look at the structure of the data.

```
data("TeenPregnancy")
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 "B","C","O","U": 3 2 2 3 4 3 4 4 2 2 ...
##  $ Church  : int  26 46 45 33 28 25 25 35 32 39 ...
##  $ Teen    : int  64 62 73 60 59 50 44 67 60 64 ...
```

FIGURE 5.15 Teen pregnancy rate versus church attendance for the 50 U.S. states

```
regmodelTPR=lm(Teen~Church,data=TeenPregnancy)
plot(Teen~Church,data=TeenPregnancy)
abline(regmodelTPR)
text(49,37,"UT")
```



Create a variable grouping church attendance and assign the median for each group.

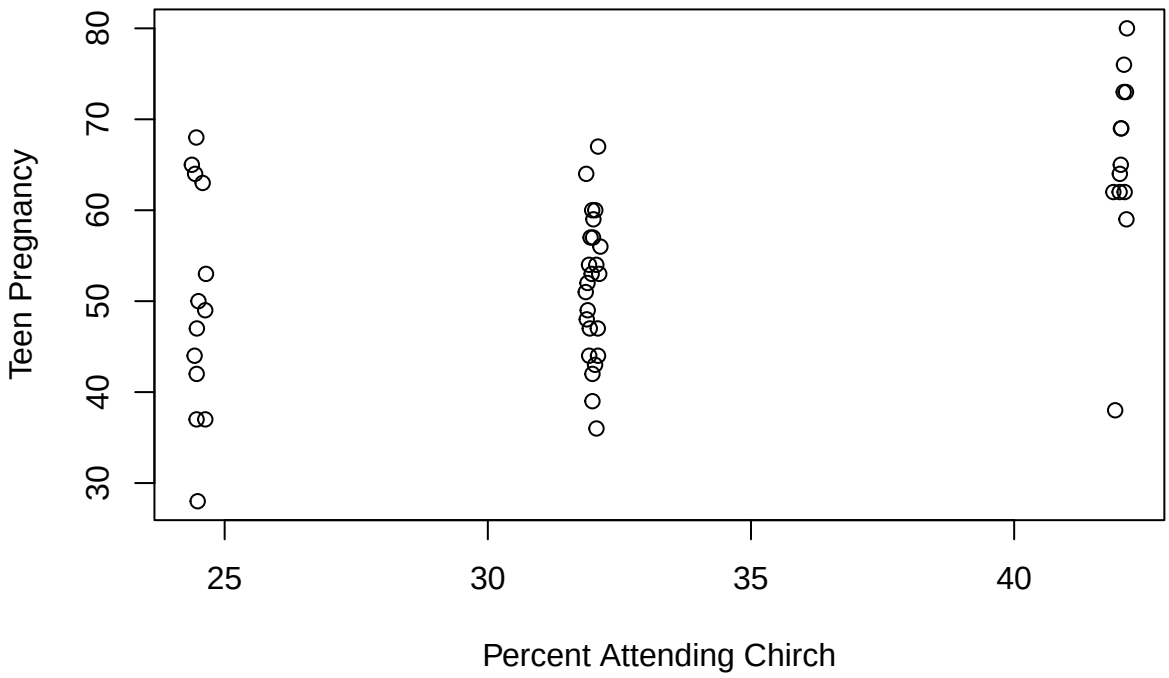
```
TeenPregnancy$ChurchGroup=32
TeenPregnancy$ChurchGroup[TeenPregnancy$Church<28]=24.5
TeenPregnancy$ChurchGroup[TeenPregnancy$Church>35]=42
```

FIGURE 5.16 Teen pregnancy rate versus church attendance

(a) Dotplots for each church group

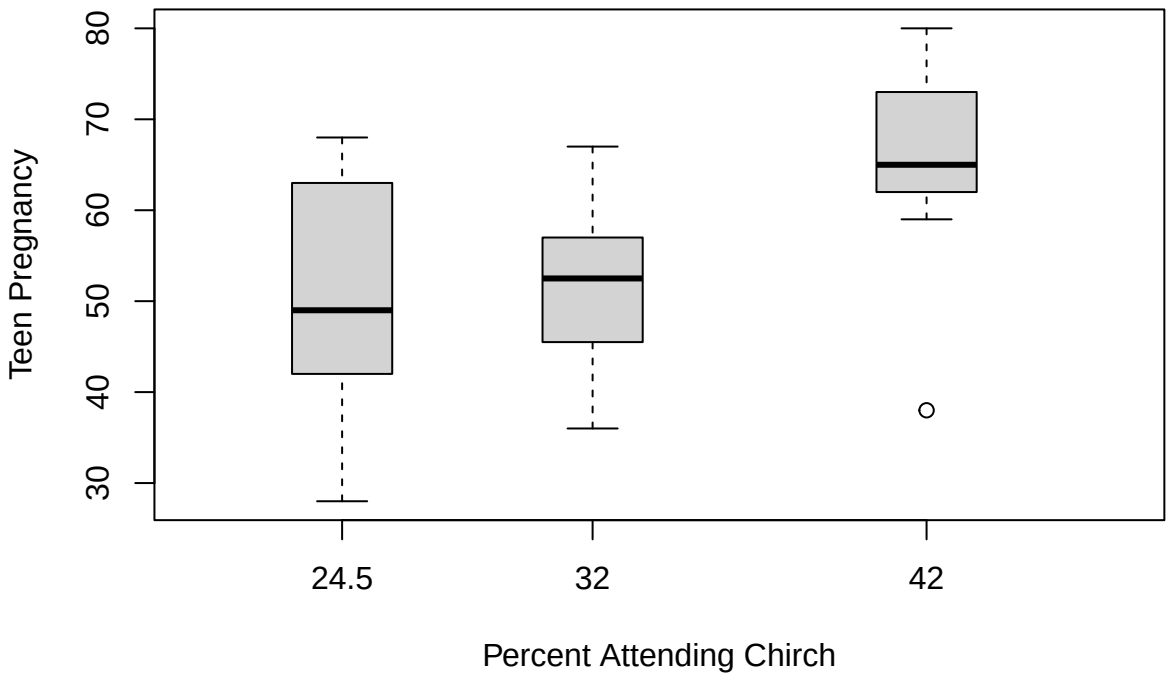
Note: The jitter( ) function randomly shifts points so they aren't at on top of each other

```
plot(Teen~jitter(ChurchGroup,0.1),data=TeenPregnancy,xlab="Percent Attending Chirch",ylab="Teen Pregnan
```



(b) Boxplots by church groups

```
boxplot(Teen~ChurchGroup,data=TeenPregnancy,at=c(24.5,32,42), xlim=c(20,48), boxwex=3,xlab="Percent Att
```


Teen Pregnancy
30
40
50
60
70
80
24.5
32
42
Percent Attending Chirch