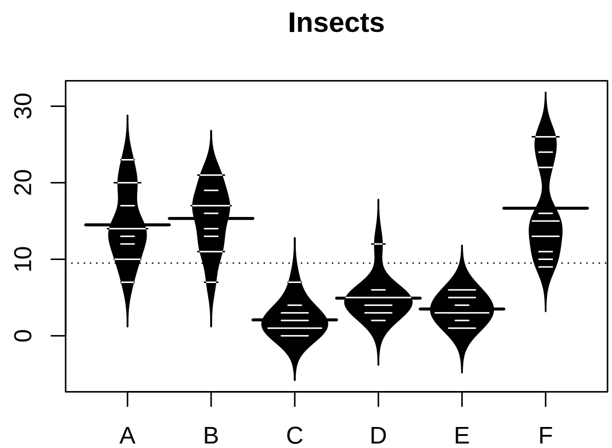
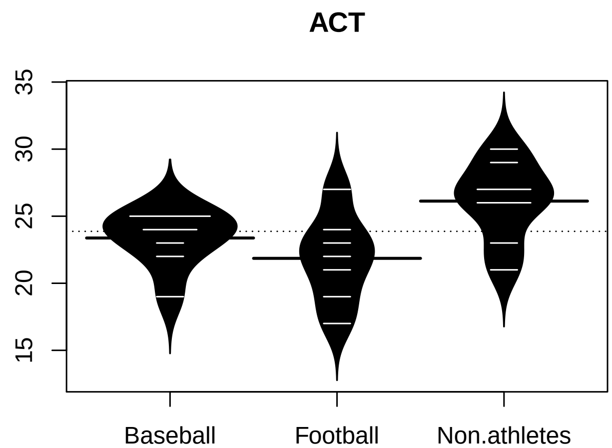
One-Way ANOVA Activity

Stat 217

In this activity we will look at two data sets: one on insecticides and the other about ACT scores.

Insects: In an agricultural experiment, 72 plots of land were randomly assigned to be treated with one of six different insecticides.The next day, the number of insects in each plot were counted. We will test to see if there is a difference in the number of insects for at least one of the sprays. Since there is a quantitative response variable and a single categorical explanatory variable, a One-Way ANOVA is an appropriate procedure for these data.





ACT Scores: A school is interested in comparing ACT scores for students with baseball scholarships, football scholarships, and non-athletes. Let Group 1 be the Baseball athletes, Group 2 be the non-athletes, and Group 3 be the Football athletes.They want to know if one of the groups has a different mean ACT score than the others.

```
##          Sport min   Q1 median   Q3 max mean   sd n missing
## 1     Baseball  19 22.8   24.0 25.0  25 23.4 2.07 8       0
## 2     Football  17 20.0   22.0 23.5  27 21.9 3.29 7       0
## 3 Non.athletes  21 25.2   26.5 27.5  30 26.1 2.95 8       0
```

1. Below is the insect Sprays ANOVA Table. Let's think about where each piece comes from.

```
lm.insect <- lm(count ~ spray, data = InsectSprays)
anova(lm.insect)

Analysis of Variance Table

Response: count
          Df Sum Sq Mean Sq F value Pr(>F)
spray      5   2669     534    34.7 <2e-16
Residuals 66   1015      15
```

Here is your key:

- J = number of groups
- N = total sample size
- Spray degrees of freedom (df) = J-1
- Residual degrees of freedom (df) = N-J
- $MS_A = SS_A/$ spray df
- $MS_E = SS_E/$ Residual df
- F-statistic = MS_A/MS_E
- The F-statistic follows an F distribution with spray df numerator and residual df denominator: $\sim F(spraydf, residualdf)$

id	Df	Sum Sq	Mean Sq	F value	Pr(>F)
spray					
Residuals					

2. Fill in the sparse ACT ANOVA Table .

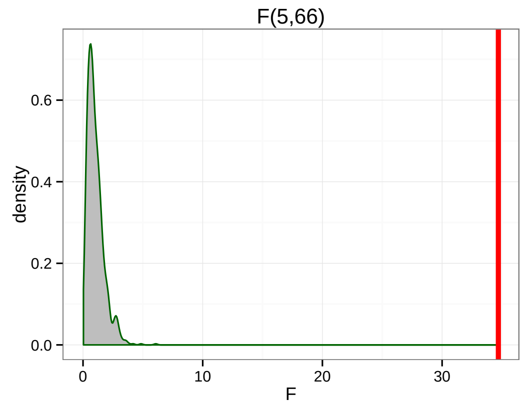
id	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group		71.0			
Residuals			7.78		

3. ACT Test Statistic Distribution

   What is the test statistic? What distribution does it follow?

4. P-value for Insect Sprays

   Recall, that we always use the right tail to find a p-value. An F(5, 66) distribution is plotted below. The test statistic is labelled with the vertical line.

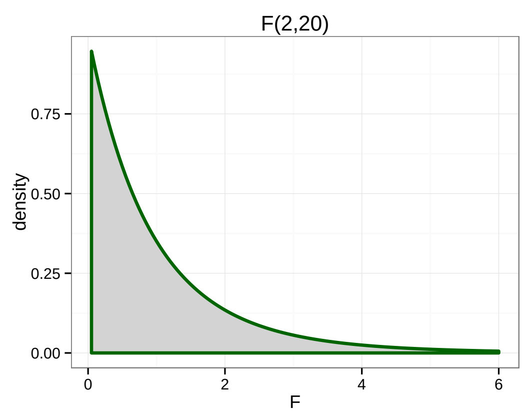


Clearly, the test statistic is more extreme than anything we would expect to observe under the null. We can use the `pf` command to compare our test statistic to the F-distribution. In R, `3.19e-17` means $3.19 * 10^{-17}$ so we get an incredibly tiny p-value for these data.

```
pf(34.7, 5, 66, lower.tail = F)

## [1] 3.19e-17
```

5. An $F(2, 20)$ is plotted below. Draw a line for the test statistic for the ACT data. Shade the part of the distribution you will use to find the p-value.



6. Write the code you would use to find the p-value.

7. What is your decision for the ACT data? Recall that the significance level was 0.05 and that a decision is either *Reject* or *Fail to Reject*.

8. What is your conclusion for the ACT data?