

Analysis of Variance (ANOVA)

Grinnell College

April 24, 2026

Tests so far:

- ▶ Difference in means
- ▶ Difference in proportion
- ▶ Goodness of fit
- ▶ Independence of two categorical variables

Today we are extending our set of tests to include testing the difference in means in multiple groups

ANOVA

The **Analysis of variance (ANOVA)** is a collection of statistical *models* used to analyze difference among many means

The null hypothesis is testing the difference of means between k groups

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

$$H_A : \text{at least one } \mu_i \neq \mu_j$$

But what does this have to do with variance?

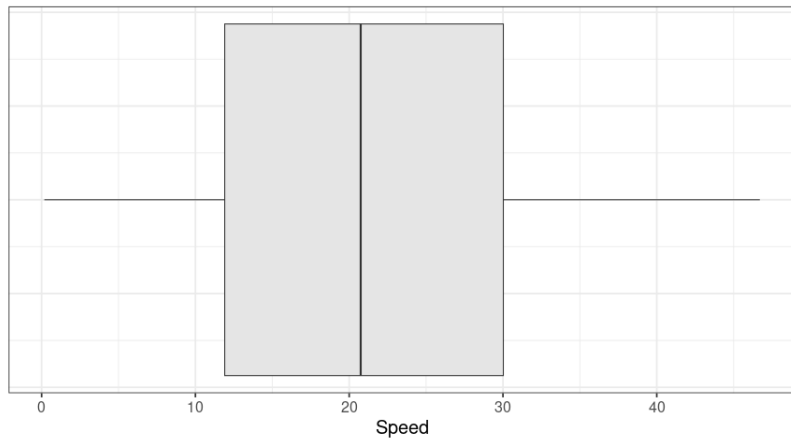
Dog Speed

Collected 400 dogs from 8 different breeds, each a sample of 50. Each set has 25 black dogs and 25 dogs of one other color. For each dog, I also recorded land speeds in miles per hour (mph)

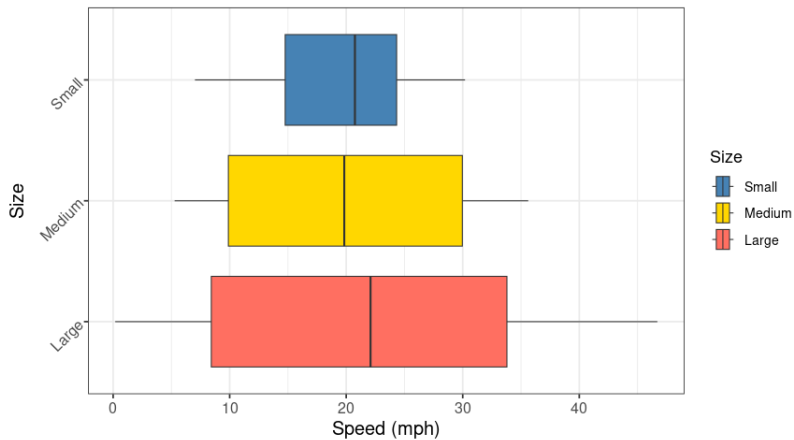
Breed	Size	Other Color	N
Chihuahua	Small	Brown	50
Corgie	Small	Yellow	50
Poodle	Medium	Brown	50
Bulldog	Medium	White	50
Saint Bernard	Large	Yellow	50
German Shepard	Large	Yellow	50
Mastiff	Large	Yellow	50
Greyhound	Large	White	50

What variables will do best in helping me predict speed?

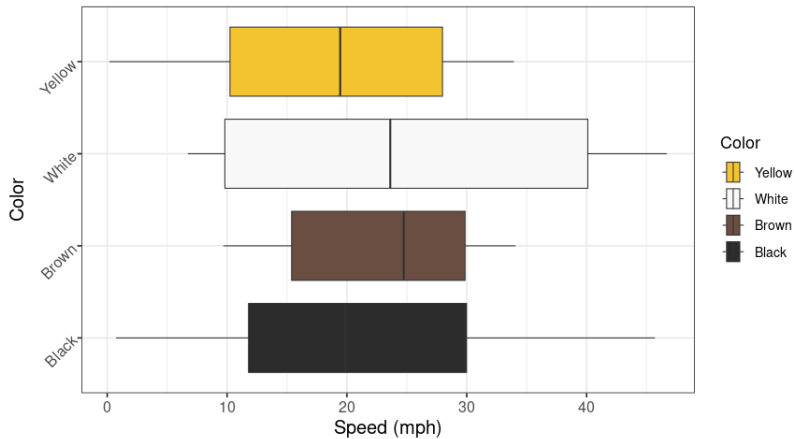
Dog Speed



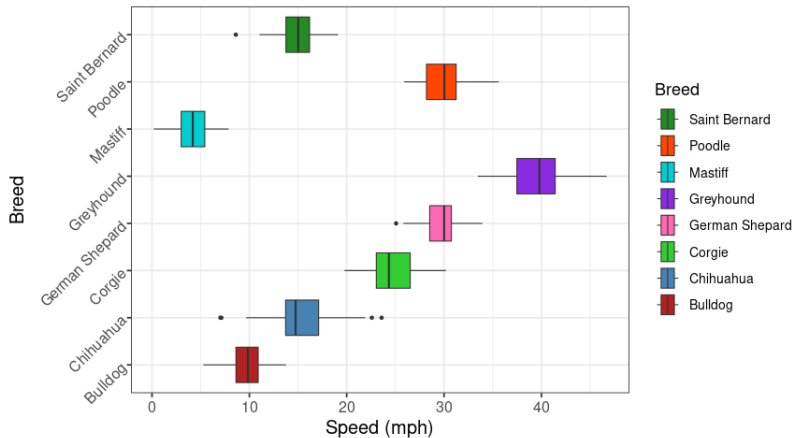
Dog Size



Dog Color



Dog Breed



The General Idea

The total variability of a sample can be broken into two parts:

- ▶ Variability within groups
- ▶ Variability between groups

How did variability *between groups* and *within groups* compare when we looked at dogs grouped by size versus by color or by breed?

Variability and You

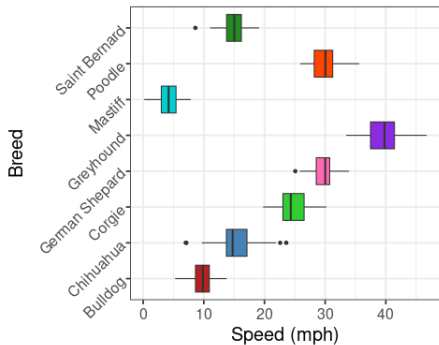
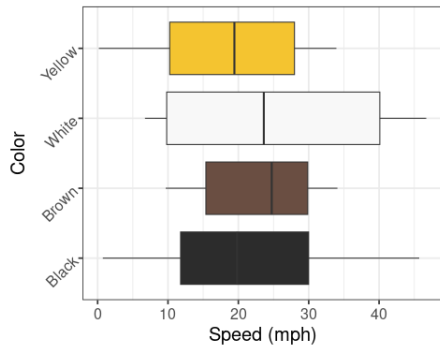
Recall our null hypothesis for ANOVA

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

$$H_A : \text{at least one } \mu_i \neq \mu_j$$

- ▶ Low within group variability \Leftrightarrow tight knit clearly defined group around a mean
- ▶ High between group variability \Leftrightarrow the groups are clearly distinct from one another

Variability and You



Variability and You (layman's terms)

What is variability

average distance from the mean

average distance of what?

why would you want to reduce it

We will think in terms of total squared error (total variability, not variance), but the average is variance

Variability and You

A common metric for variability is the sum of squares giving total squared distance between observations and mean

$$\text{TSS} = \sum_i^n (x_{ij} - \bar{x})^2$$

where $i = 1, \dots, n$ indicates the observation and $j = 1, \dots, k$ indicates the group. We can always decompose this into a sum of two sums of squares

$$\underbrace{\sum_i^n (x_{ij} - \bar{x})^2}_{\text{Total variability}} = \underbrace{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}_{\text{Variability within groups}} + \underbrace{\sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2}_{\text{Variability between groups}}$$

Within-group Variability (Layman's terms)

Within group variability gets at idea: what if I told you the group this was in. How much error left?

Compare all auto trans to manual trans. Have things gotten better?

This still represents error. If all I know is greyhound, then i still have error.

The sum of squared errors basically says, "okay, if I take each observation and guess it's group mean, how much total error is left?"

It also makes sense for me to ask about the average of this

Within-group Variability

Within-group variability is associated with a standard *sum of squared errors (SSE)* where we assume each observation i in group j is denoted x_{ij} and are compared to the mean of group j :

$$\text{SSE} = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$

This can also be written as the *weighted* sum of group variances

$$\text{SSE} = \sum_{j=1}^k (n_j - 1) \hat{\sigma}_j^2 = (n_1 - 1) \hat{\sigma}_1^2 + (n_2 - 1) \hat{\sigma}_2^2 + \cdots + (n_k - 1) \hat{\sigma}_k^2$$

As this sum involves estimating k different group means, \bar{x}_j , we have $n - k$ degrees of freedom, giving the mean of this metric to be

$$\text{MSE} = \frac{\text{SSE}}{n - k}$$

Variability between groups (layman's terms)

Between group variability gets at the idea: how different are my groups from the overall average?

How far is average greyhound from average dog? What about average corgie?

If on average our groups are very different, this is good – telling you that dog is greyhound very informative vs telling you if it corgie

If average of groups are all pretty similar (and near the overall mean), knowing that group information doesn't give you much better of a guess.

Variability between groups

This describes how different each of the groups are from one another

Also known as the sum of squares between groups (SSG), we can compute it by finding the weighted mean of group deviations:

$$\begin{aligned}SSG &= \sum n_i(\bar{x}_i - \bar{x})^2 \\ &= n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_k(\bar{x}_k - \bar{x})^2\end{aligned}$$

As there are k groups, we can find the mean by dividing by k , less 1 degree of freedom from finding \bar{x} ,

$$MSG = \frac{SSG}{k - 1}$$

F-statistic (layman's terms)

The statistic we are going to consider is the average between group variability and the average within group variability.

$$F = \frac{\text{Average variability between groups}}{\text{Average variability within groups}}$$

When the null hypothesis is true, the sampling distribution of this statistic will be an F distribution

The simplest way to understand is to realize that:

1. If my null hypothesis is true
2. Then all of the group means in my sample will be similar
3. Then the variability *between my groups* will be small
4. Then the numerator of my F statistic will be small
5. Then all things equal, a big F is more evidence against my null hypothesis

F-statistic Ratio

Finding the means of SSG and SSE help keep the metrics interpretable when the number of groups or samples increases

Ultimately, then, what determines value the outcome of our test is the ratio between group variations and variation from error

$$F = \frac{MSG}{MSE}$$

What makes the F statistic larger:

- ▶ MSG increases
- ▶ MSE decreases

F distribution

Just as we are able to use the t -distribution in finding p -values for the difference of two means, we can use the F distribution to find a p -value for assessing the null hypothesis for ANOVA

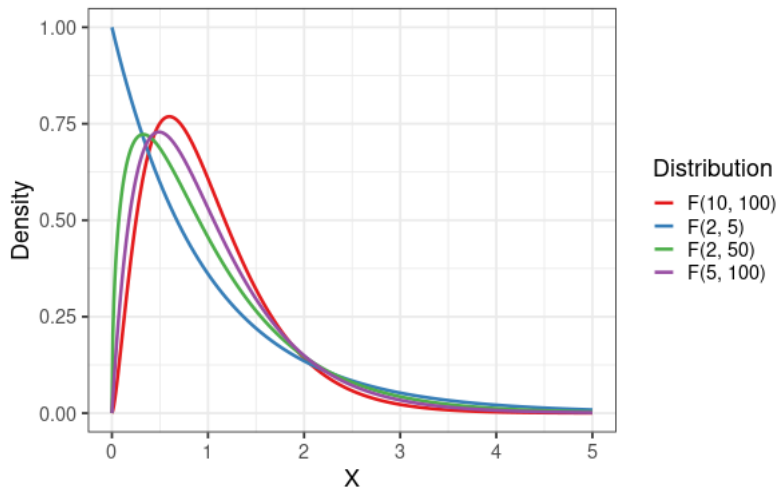
Generally speaking, we are in good shape if:

- ▶ The distributions of the groups are roughly normal
- ▶ The variances between the groups are roughly similar. Generally so long as the standard deviation of one group doesn't exceed twice that of another

Again similar to the t -distribution, the F distribution is associated with degrees of freedom, in this case two, one for each of the mean squares in the ratio.

F distribution

$$F = \frac{MSG}{MSE}$$



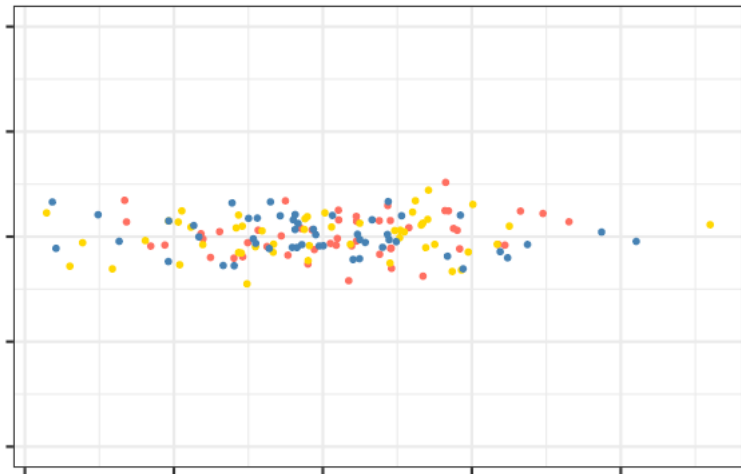
Formulas

$$\underbrace{\sum_i^n (x_{ij} - \bar{x})^2}_{\text{SST}} = \underbrace{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}_{\text{SSE}} + \underbrace{\sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2}_{\text{SSG}}$$

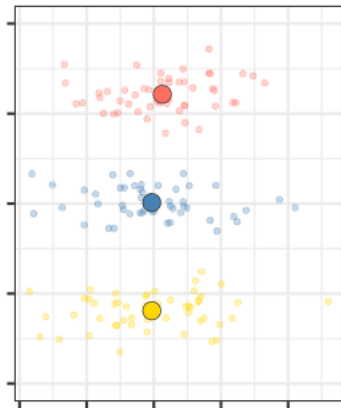
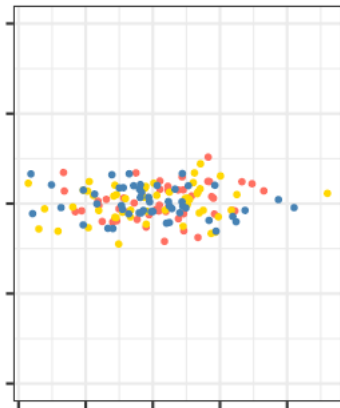
- ▶ SST = SSE + SSG
- ▶ SSE = sum of squares *within groups*
- ▶ SSG = sum of squares *between groups*
- ▶ $MSG = \frac{SSG}{k-1}$
- ▶ $MSE = \frac{SSE}{n-k}$
- ▶ $F = \frac{MSG}{MSE} = \left(\frac{n-k}{k-1} \right) \frac{SSG}{SSE}$

Source	df	Sum Sq	Mean Sq	F value	Pr(>F) / p-value
Group	k-1	SSG	$MSG = \frac{SSG}{k-1}$	$F = \frac{MSG}{MSE}$	Upper tail
Error	n-k	SSE	$MSE = \frac{SSE}{n-k}$		
Total	n - 1	SSTotal			

Example 1

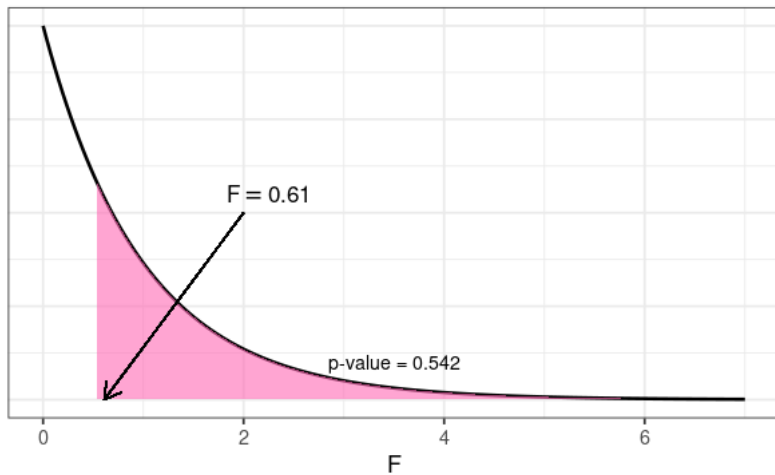


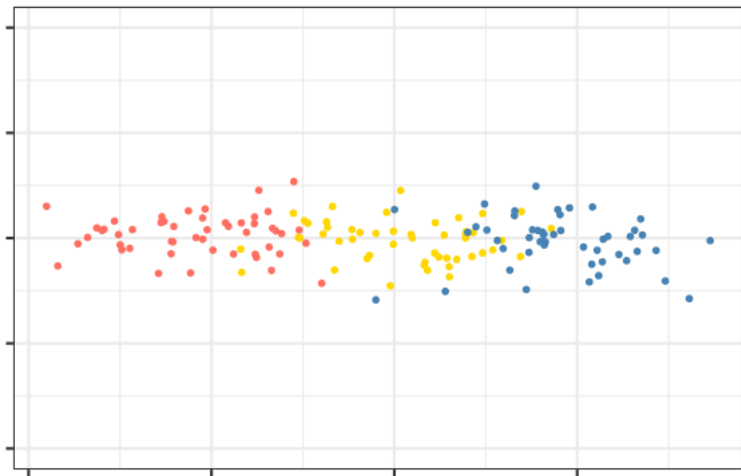
Example 1



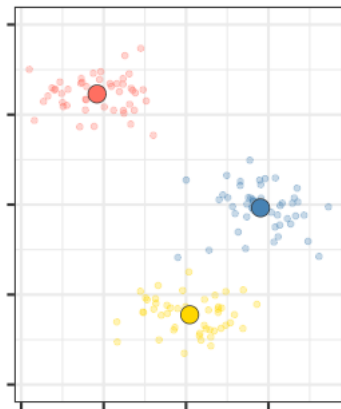
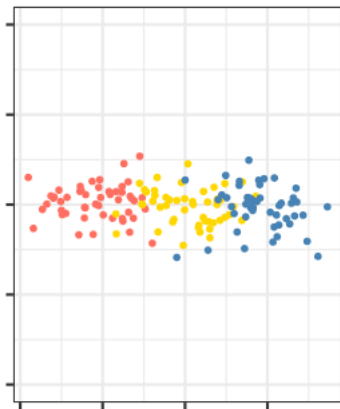
	Df	Sum Sq	Mean Sq	F value	<i>p</i> -value
Group	2	5.09	2.54	0.61	0.5427
Residuals	147	609.07	4.14		

F(2, 147) Distribution



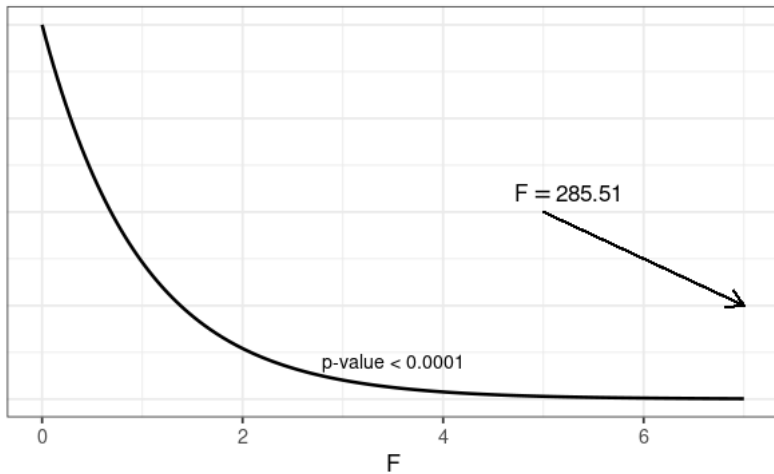


Example 2

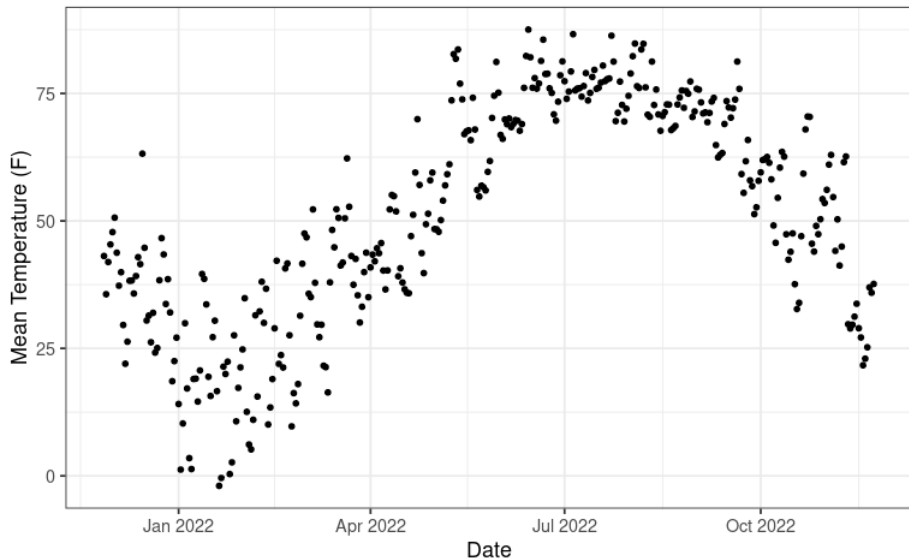


	Df	Sum Sq	Mean Sq	F value	<i>p</i> -value
Group	2	2587.33	1293.67	285.51	<0.00001
Residuals	147	666.07	4.53		

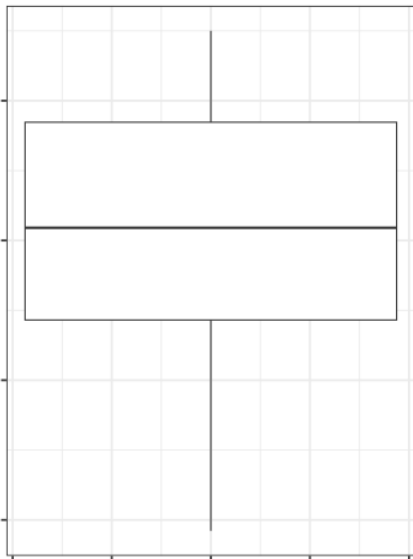
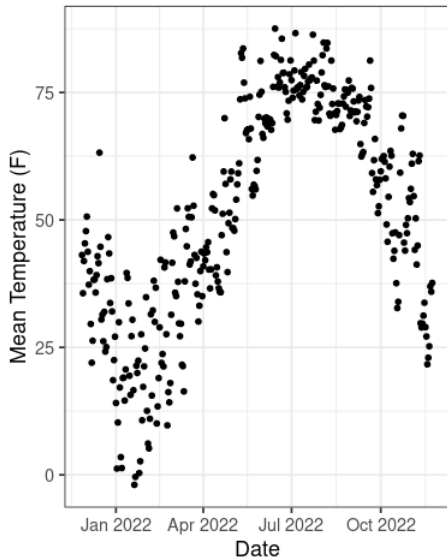
F(2, 147) Distribution



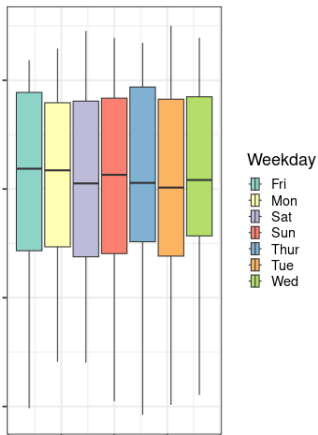
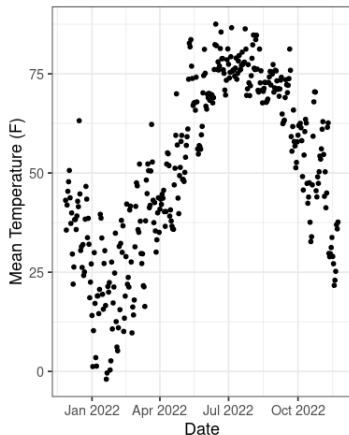
Annual Temperature – Grinnell



Annual Temperature

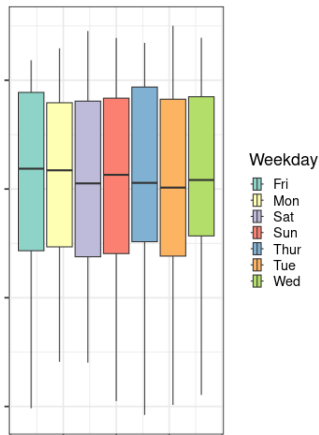
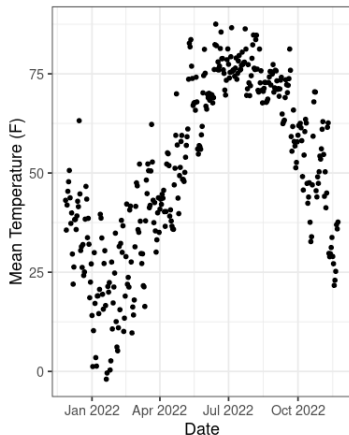


Temperature by Day



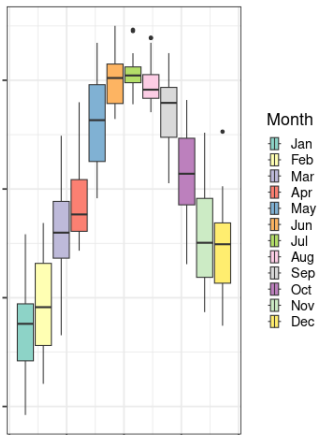
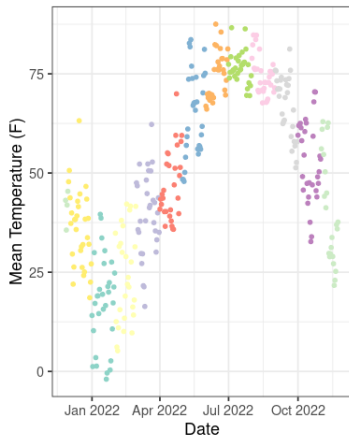
	Df	Sum Sq	Mean Sq	F value	p-value
Weekday					
Residuals					

Temperature by Day



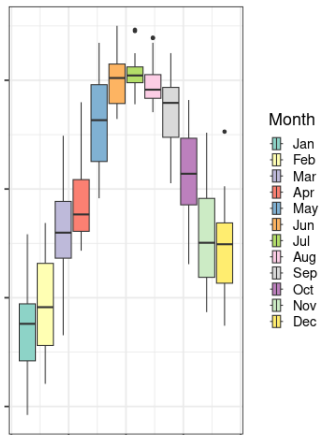
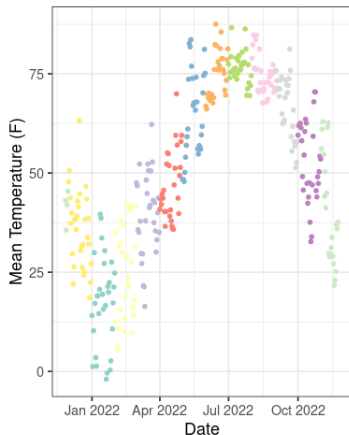
	Df	Sum Sq	Mean Sq	F value	<i>p</i> -value
Weekday	6	342.71	57.12	0.12	0.9939
Residuals	355	168524.83	474.72		

Temperature by Month



	Df	Sum Sq	Mean Sq	F value	p-value
Weekday					
Residuals					

Temperature by Month



	Df	Sum Sq	Mean Sq	F value	<i>p</i> -value
Month	11	138048.06	12549.82	142.52	<0.0001
Residuals	350	30819.48	88.06		

Review

- ▶ ANOVA used to test equality of means between groups
- ▶ Total variability made up of within and between group variation
- ▶ ANOVA gives us a method for determining the relative sizes of these types of variation
- ▶ F statistic can be used as a measure of certainty
- ▶ Larger F statistic leads us to reject H_0