Lecture 2

Scatterplots

Last semester, we briefly discussed scatterplots as a way of visualizing the relationship between two numerical variables. We did very little with it last semester, but this semester, we will rely on it quite heavily.

For instance, if we wanted to look at how mpg from our mtcars dataset related to the disp(lacement) variable, we can plot them against each other in a scatterplot.
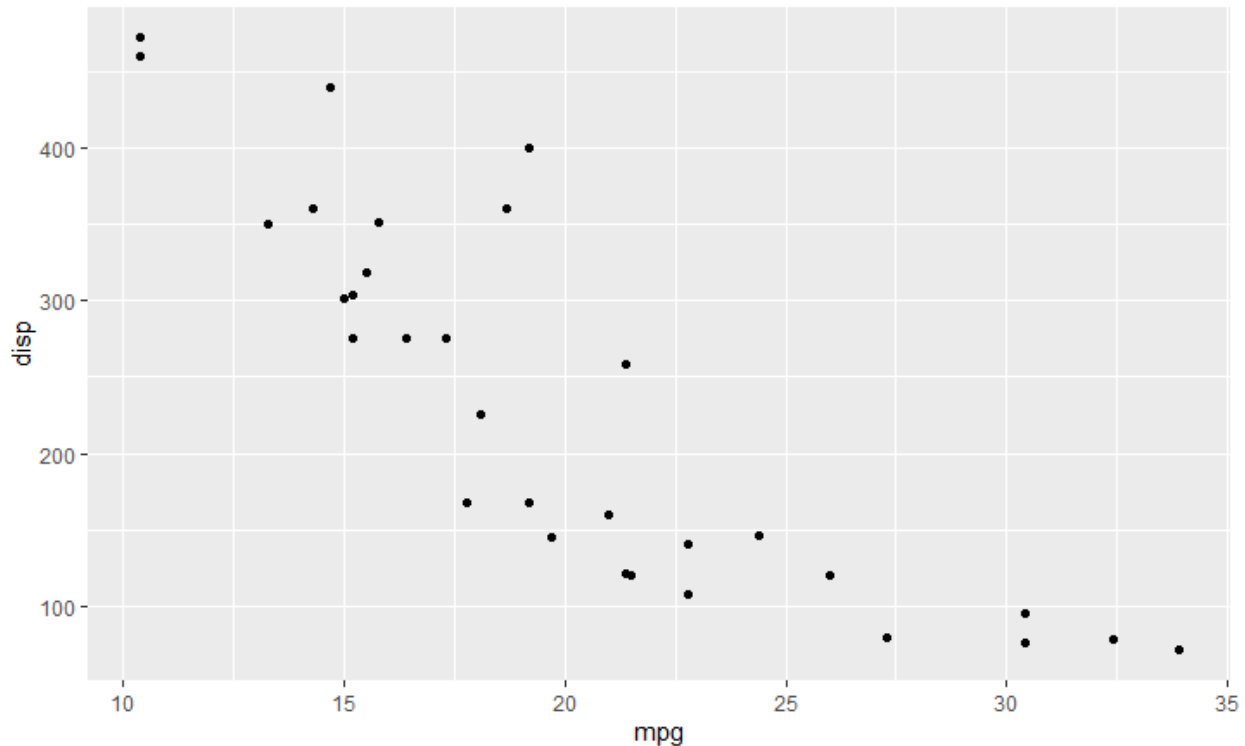

In this configuration, we usually think of the independent variable on the horizontal axis, and the dependent variable on the vertical axis. The correlation will not depend on that orientation, but the way that we think of the causal relationship and our regression equation will make this choice matter. For now, we can say that mpg and disp have a negative correlation, that may not be purely linear. In general, as mpg increases, the disp variable decreases (an vice versa).

Correlation

We want to consider the relationship between the value of a correlation and how that looks in a scatterplot. We can use R to simulate different correlation values. The gallery of graphs below shows you different correlation values from $r = -0.99$ through $r = 0.99$. (Perfect correlations of ± 1 fall on a perfectly straight line.)

As you can see from the graphs, negative correlation means that the linear relationship has a negative slope. A positive correlation has a positive linear slope. As the correlation tends to 0, the slope of the line tends to the horizontal. The points also spread out further and further from the line that best fits through the middle of the data.
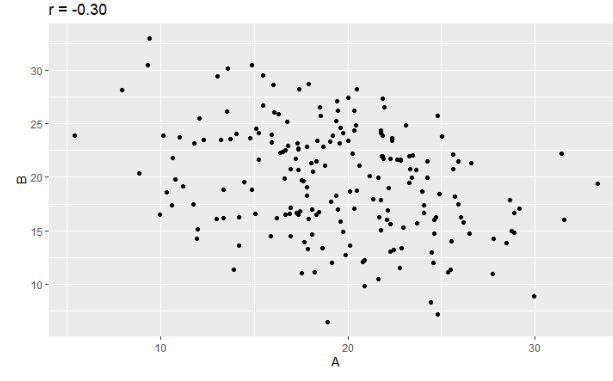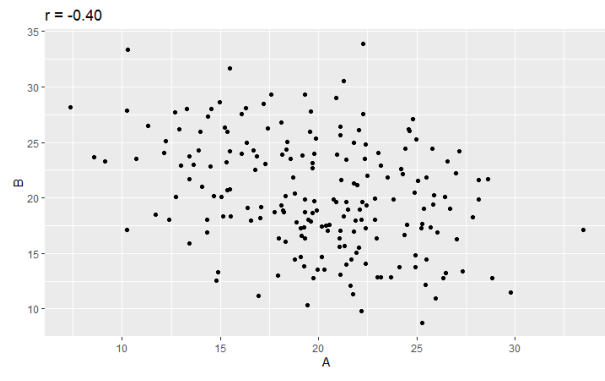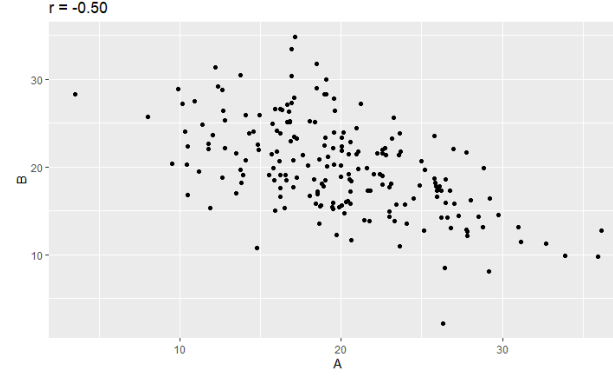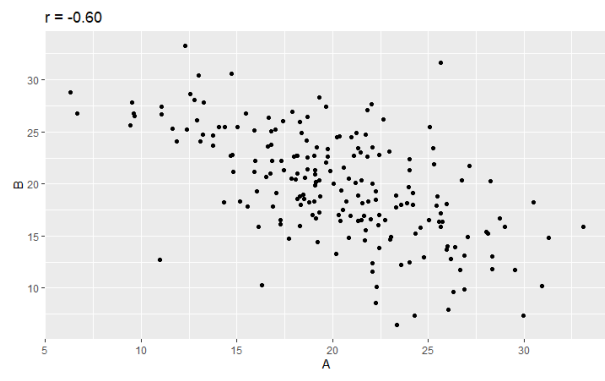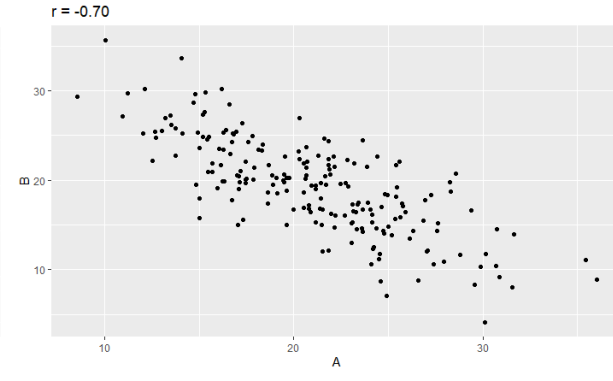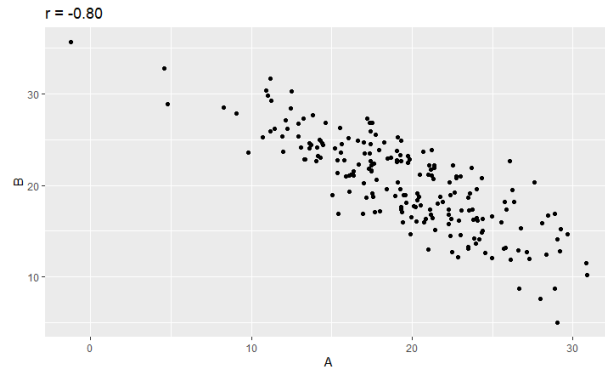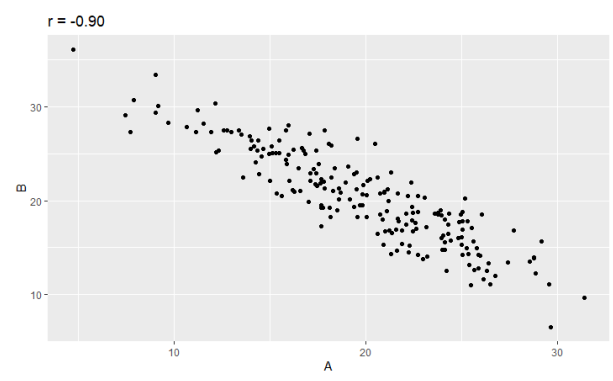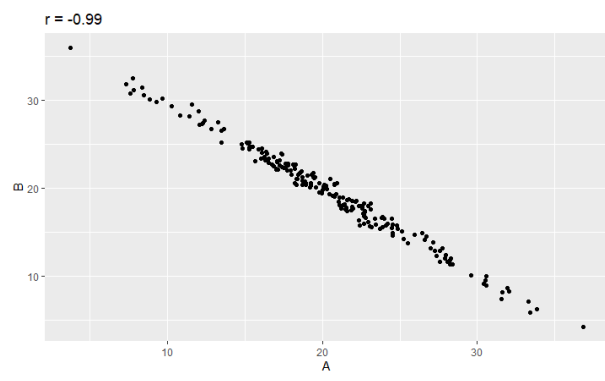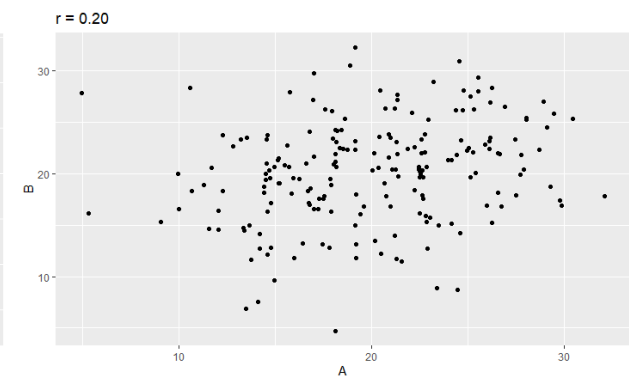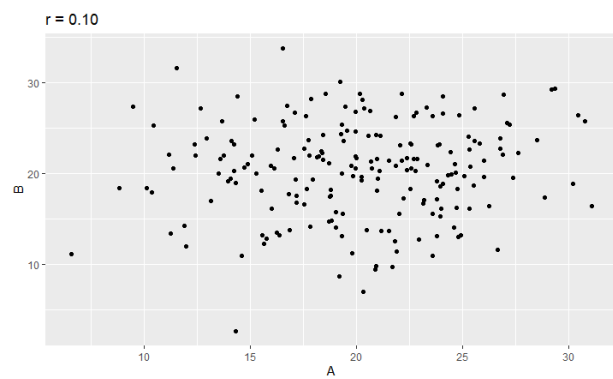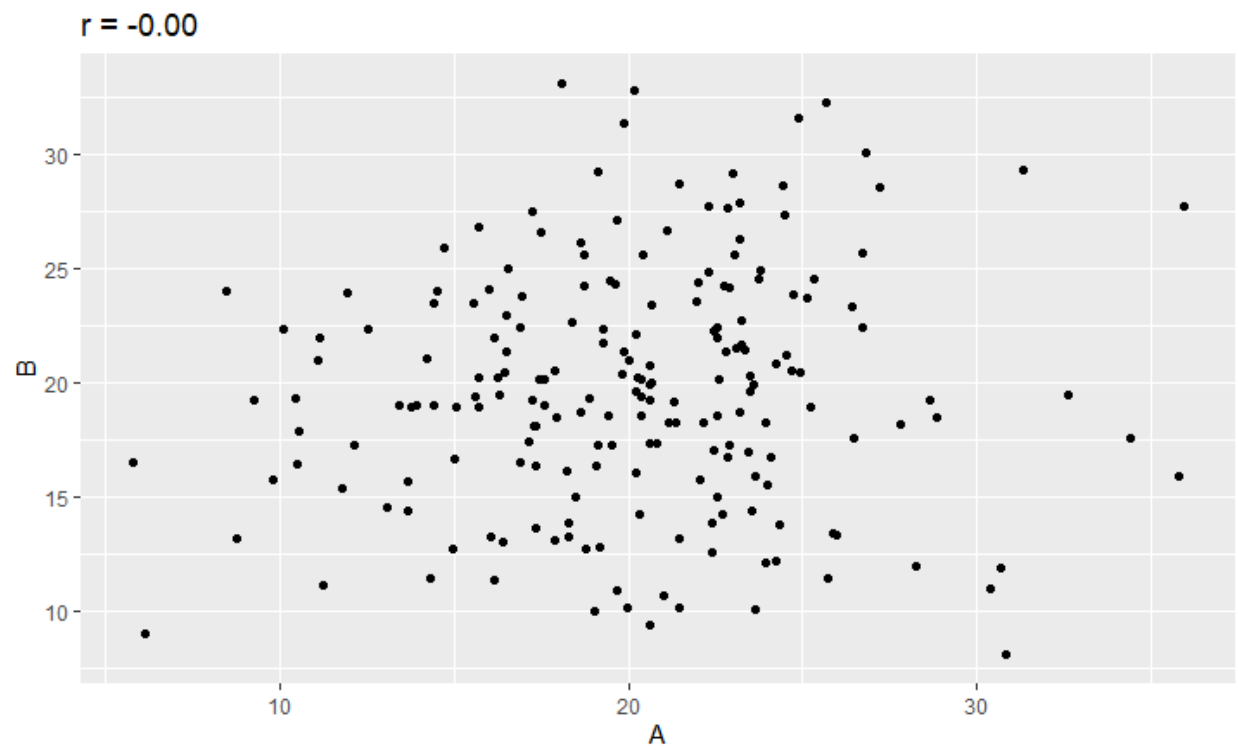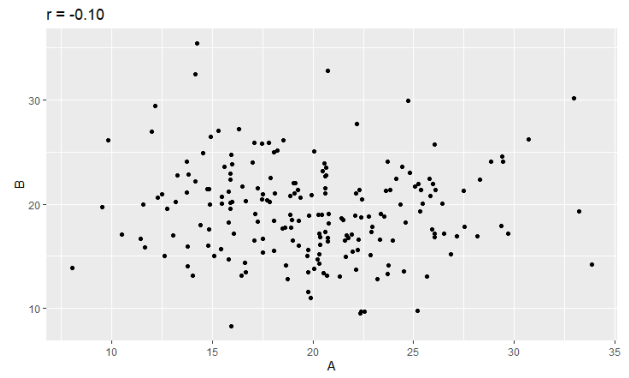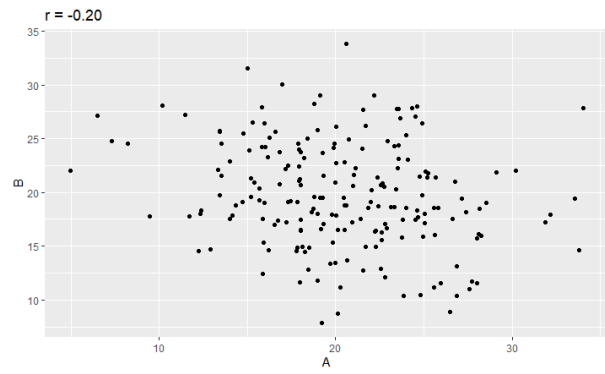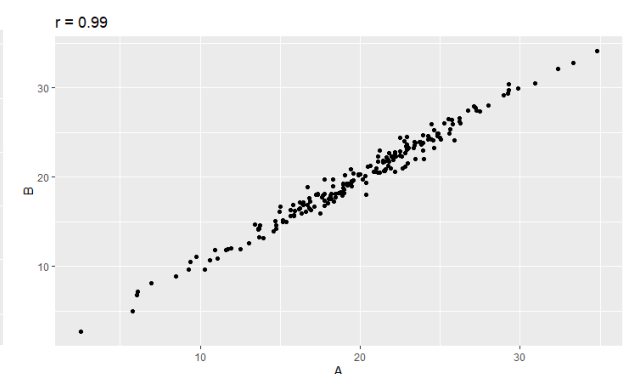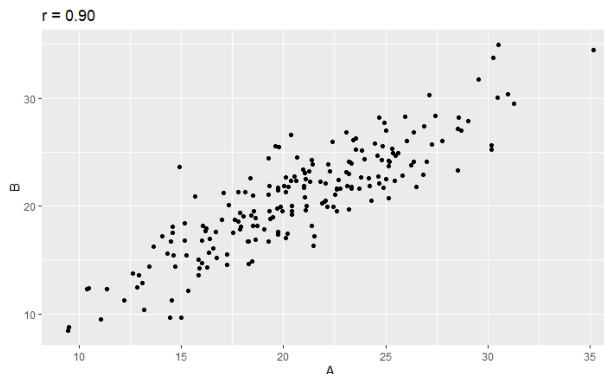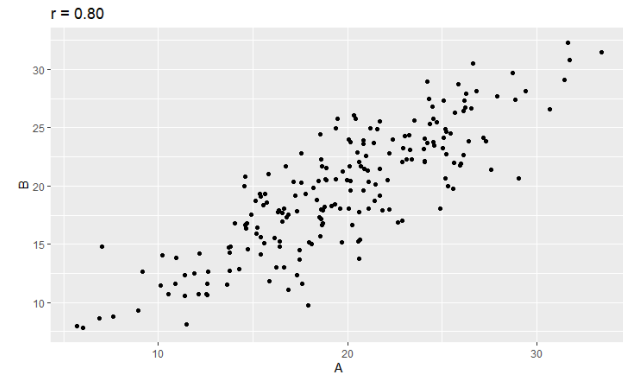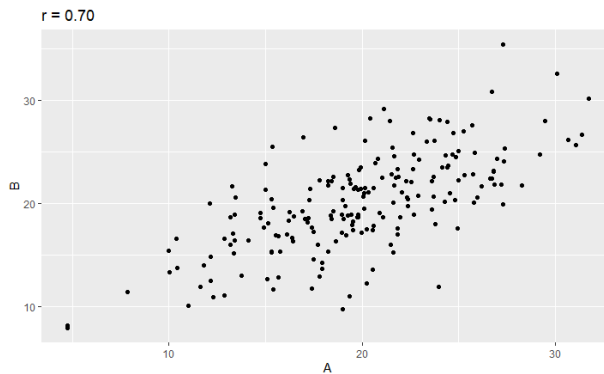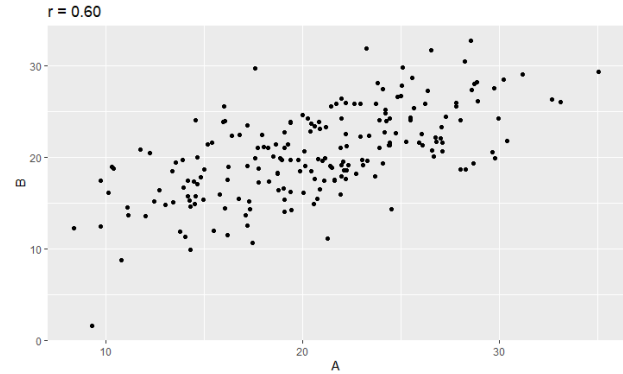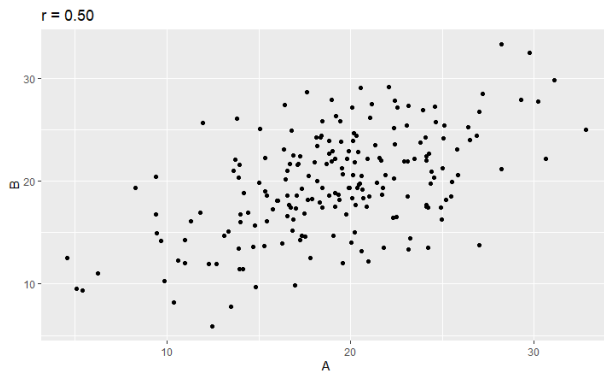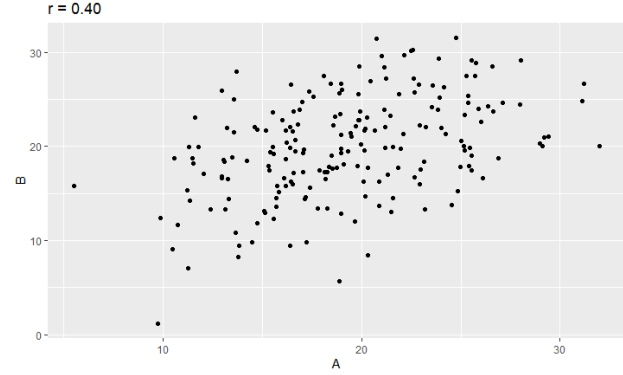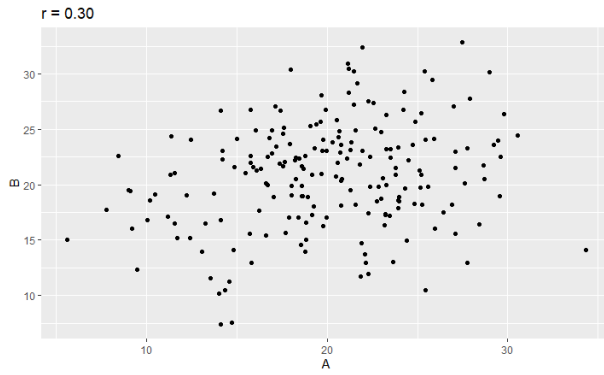
Correlation values are sometimes grouped into broad levels of strong, moderate and weak. Strong is considered $|r| > 0.7$ (some texts use 0.8). this corresponds to the first two and last two rows of our example scatterplots. Weak correlation is $|r| < 0.4$. The middle large plot is 0, and the two rows above and below that are weak correlations. You can barely tell which direction the line is even pointed in the

weakest cases. Moderate is in between those values, roughly $0.4 < |r| < 0.7$. Values near the boundaries are sometimes described as moderately weak, or moderately strong, but there are no set boundaries for these terms.

A fun way to practice and develop your correlation intuition is with the Correlation Game: http://guessthecorrelation.com/

All the correlation examples in our sample scatterplots have a linear relationship. We are going to start with this simple scenario, but we also need to know when not to use this. Later in the course we'll address non-linear cases, but how do we know when it's okay and when it's not. Let's look at some non-linear cases that we should not be modeling with our linear models.

Linear or Nonlinear Relationships
Let's start with the linear relationship for comparison.
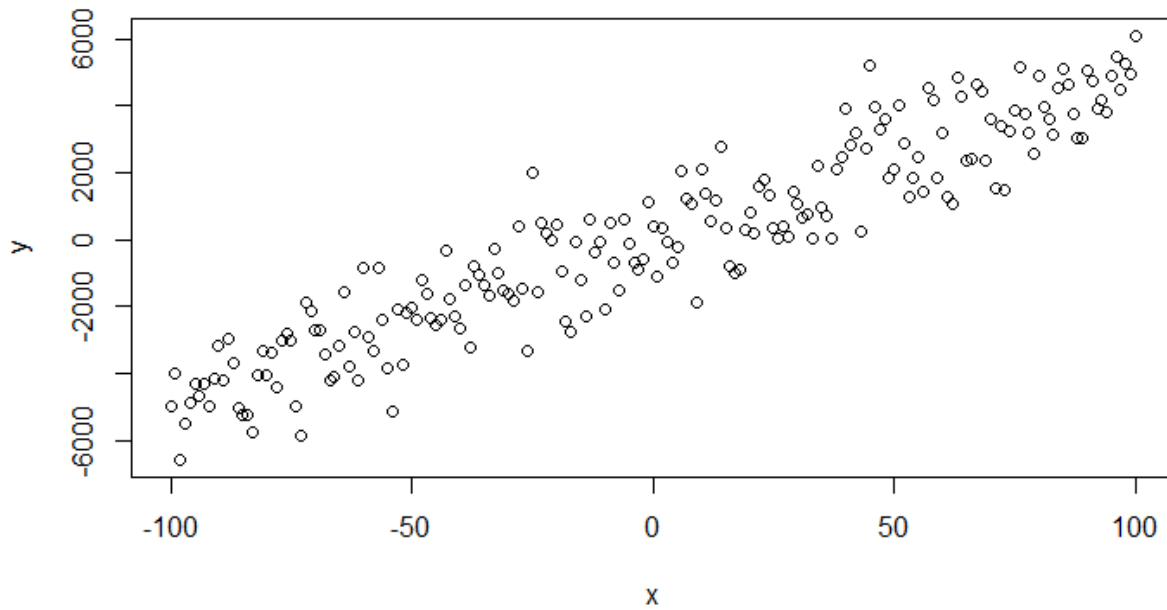
What would some nonlinear cases look like?
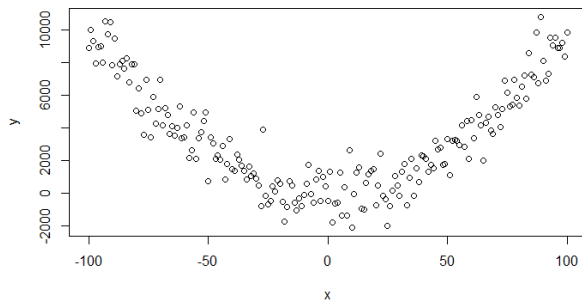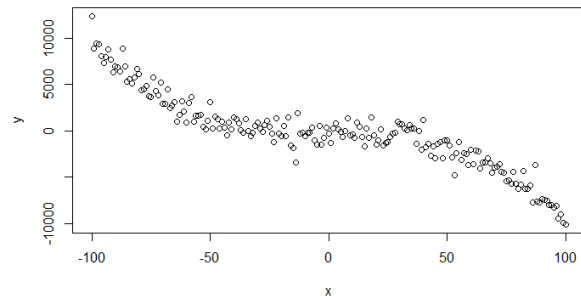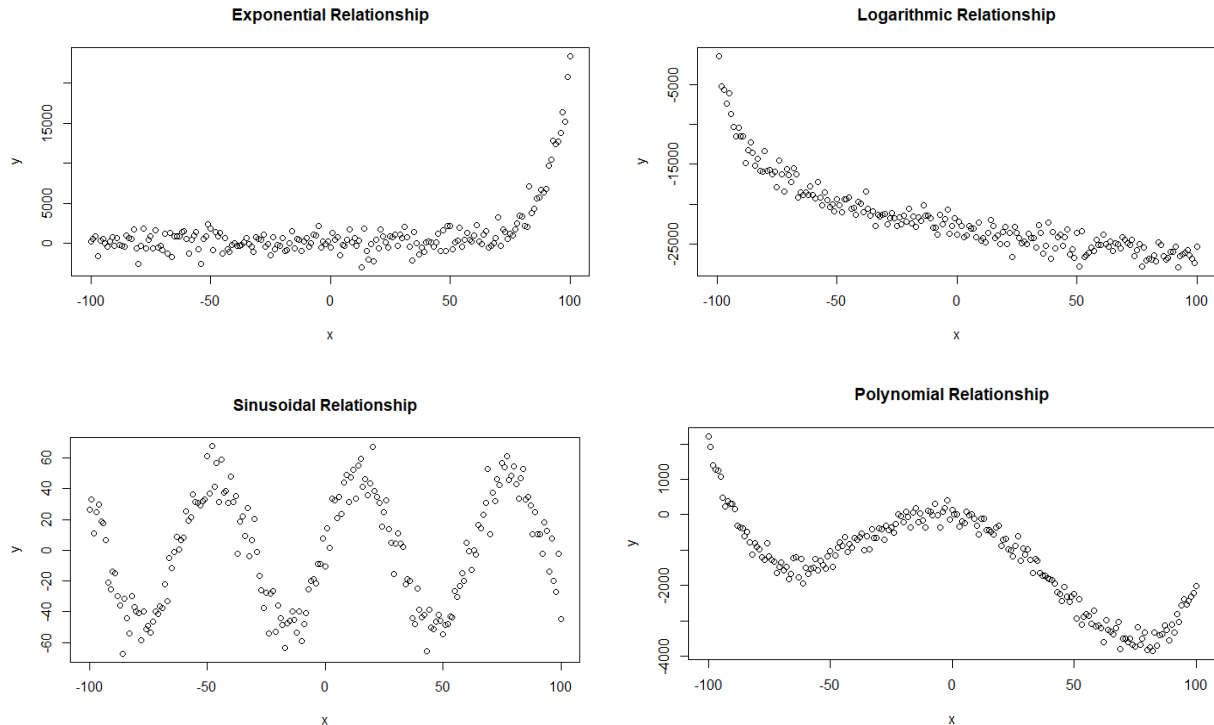
Sometimes it can be difficult to tell which degree of polynomial is required, or whether a graph is a power model, an exponential model or a log model, but for the moment, the important distinction is that none of these graphs are linear. We'll return to the non-linear case later on. Some of these non-linear models can be obtained from a transformation of one or more variables, but not all.

Pearson correlation for linear models is also connected to a more general concept we will discuss in greater depth a couple lectures from now: the coefficient of determination. This measurement is typically given the value R^2. All model types have a coefficient of determination, even ones that do not have a linear correlation, but linear models that have a linear correlation value, one can square it to obtain the coefficient of determination, i.e. $r^2 = R^2$. Since the values of r are between -1 and 1, the values of R^2 are between 0 and 1. As we'll see, the coefficient of determination measures how much we can reduce the variability in the prediction variable y by knowing it's relationship to the predictor variable x.

If you've ever worked with Excel to do regression, you may recall that Excel graphs can display the coefficient of determination on the graphs. For linear models, you can back out of that value to find the correlation: take the square root. The sign of the correlation will match the sign of the slope of the linear model.

Other Measures of Correlation
In addition to the Pearson correlation, there are other method for measuring something like correlation using distribution-free methods with fewer assumptions.

Spearman's correlation measures whether a relationship is monotonic or not. Monotonic means it's always increasing or always decreasing. This will cover linear cases, but also some of the non-linear cases, such as the cubic example above, the exponential and logarithmic models, but not the sine,

quadratic or general polynomial model. Like Pearson's correlation, the values also fall between -1 and 1. In Pearson's correlation, the variables need to be numerical (interval or ratio level), but with Spearman's correlation we can also include ordinal level variables. The variables do not need to be normally distributed and neither do the errors, which is why it is considered non-parametric or distribution-free.

The general procedure to calculate Spearman's correlation is to rank the x-values and the y-values, and then plot the ranks of x against the ranks of y, and then calculate the Pearson correlation on these pairs of values. If the values are well-spaced, the noise won't be large enough to cause y-values to go against the general increasing or decreasing trend, which will result in a Spearman's correlation of -1 or 1, even if the curve is not a straight line.

$r_s = -1$
perfect −ve
monotonic correlation

$r_s = 0$
no correlation

$r_s = 1$
perfect +ve
monotonic correlation
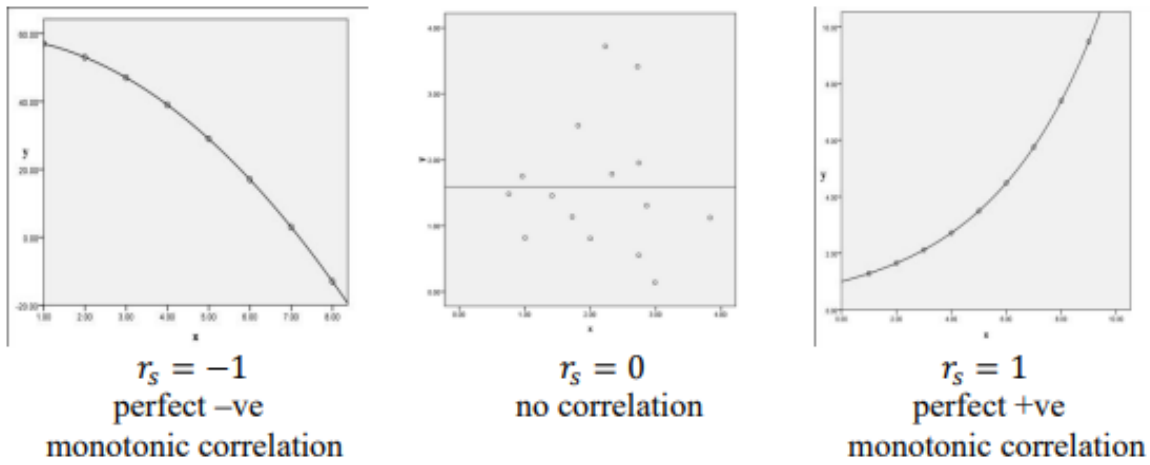
As with Pearson's correlation, curves that are strongly nonmonotonic will also have Spearman's correlation of zero, such as our quadratic or sine curve.

To differentiate Spearman's correlation from Pearson's correlation, we may notate the variable with a subscript, r_s or ρ_s. We can conduct hypothesis tests on the Spearman's correlation as well. We'll return to this later on and we can let R do the work for us.

We can also use another correlation measure called Kendall's tau or Kendall's tau-b. As with Spearman's correlation, Kendall's tau is based on ranks and so is nonparametric. It produces values between 0 and 1 (if get negative values, this is only due to the order of subtraction, and so you can ignore it). As with other measures of correlations, values close to 1 are strong (1 is perfect) and 0 is no correlation.

There are several versions of Kendall's tau that depend on factors like whether there are ties, but generally it can be calculated by

$$\tau = \frac{C - D}{C + D}$$

Where C is the number of concordant pairs and D is the number of discordant pairs. After ranking the pairs of values, determine the concordant pairs by how many larger ranks are below a certain rank. Discordant pairs work similarly except that you are looking for ranks that are smaller instead of ranks that are bigger. Then sum the values to obtain C and D respectively.

R can compute these values for us, so we'll be able to experiment with them in the lab.

Linear Algebra
In data science, we often don't use the algebraic or summation formulations to solve for things that could be solved using linear algebra. There is no way we can cover all the theory required to discuss this in depth, but we are going to do a quick crash course in the basics of linear algebra in order to be able to discuss some of the main concepts we'll encounter later on.

Linear algebra is based on matrices and vectors (we can think of vectors as just a special case of matrices). Let's discuss some basics of matrix properties and operations.

In simplest terms, matrices are rectangular arrays of numbers. One of the most common usages for them is as a means of obtaining solutions to (linear) systems of equations. Computers like working with arrays of numbers much more than they like working with equations. They also allow us to make generalizations about solutions to systems of equations that are independent of the size of the system. As we will see, it is also possible with matrices to solve for a single variable of a system without solving the whole thing.

I. **General Properties**
    A. **Size of Matrices**
        Matrix sizes are described in two dimensions: rows × columns.

$$\begin{array}{c} col1 \quad col2 \\ \begin{array}{c} row1 \\ row2 \\ row3 \end{array} \begin{bmatrix} -1 & 4 \\ 3 & -5 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \end{array}$$

So, the matrix in the above example is describes as a 3×2 matrix. When entries in the matrix are numbered, they are likewise described in the row x column format. The entry in the first row first column is a_{11}. The entry in the third row second column is a_{32}.

Matrices can be any dimension bigger than zero, from 1×1 to any $m \times n$. Some examples of matrices are below.

$$[5], \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 5 \\ 8 \\ -9 \end{bmatrix}, \begin{bmatrix} 0 & -2 & -4 & 3 \\ 5 & 1 & 0 & 0 \end{bmatrix}, \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

Here, we have a 1×1, 2×2, 5×1, 4×2 and an arbitrary $m \times n$.

    B. **Matrix Notation**
        Capital letters are used to refer to matrices: **A**, **I**, etc. The only letter used to describe a matrix is the zero matrix, which is represented by a **0**. Matrices are often boldfaced in typed text. Occasionally, matrices will be represented by a single entry [a_{ij}], with i ranging from 1 to m and j ranging from 1 to n.

Letters with paired subscripts, as shown above, always represent entries of a matrix. To differentiate between a single entry and the matrix represented by a general entry, the square brackets [] go around the matrix, while the individual entry lacks the brackets.

To represent a row of matrix subscripted R is used: R_1, R_2, etc. Similarly for columns: C_1, C_2, etc.

Notation for operations on matrices will be address below in Section II.

C. Special Matrices

a. *Identity Matrix*

The identity matrix **I** is a square matrix (size $n \times n$) with 1's along the diagonal and zeros in all other entries. When multiplied by another matrix of suitable size (see Section IIC below), **AI=A** the original matrix is result. The identity matrix is also the result of multiplying a matrix by its inverse (see Section IIE below). **AA^{-1}=I**.

Below are examples of identity matrices of various sizes.

$$[1], \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

b. *Zero Matrix*

The zero matrix **0** stands in for the role of zero in matrix land. The zero matrix, like the identity matrix, is actually a set of matrices of appropriate size, where all the entries are zero. Under matrix multiplication, the result is always zero. **A0=0**.

$$[0], \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Unlike the identity matrix, the zero matrix need not be square. It just needs to be whatever size is necessary to make the multiplication work.

c. *Triangular Matrix*

A triangular matrix is a matrix with entries that are potentially non-zero on the diagonal and potentially on one side of it, but are zero on the other side. Triangular matrices can be divided into two classes: *upper triangular* if the entries below the diagonal are zero (i.e. the nonzero entries are in the upper half), and *lower triangular* if the entries above the diagonal are zero (i.e. the nonzero entries are in the lower half).

Triangular matrices are useful because they make multiplication easier, and can be used to help solve systems of equations. Triangular matrices must be square matrices (see Section e below).

$$[1], \begin{bmatrix} -1 & 3 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 4 & 2 & 0 \\ 6 & -3 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 5 & 0 & -8 \\ 0 & -1 & 0 & 7 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix}$$

d. *Diagonal Matrix*

A diagonal matrix is a square matrix (see Section e) with nonzero entries only on the diagonal. An identity matrix is a space case of a diagonal matrix with only ones on the diagonal.

$$[2], \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 4 \end{bmatrix}, \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & -7 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 11 \end{bmatrix}$$

Diagonal matrices make it possible to do rapid matrix multiplication, and can be used to solve systems of equations. Calculating operations like inverses and determinants are also made quite simple.

e. *Square Matrix*

A square matrix is a matrix with the same row and column dimensions: 1×1, 2×2, 3×3, etc. Certain operations like inverse and determinants are only defined on square matrices. As we've seen above, identity matrices, triangular and diagonal matrices are all required to be square.

f. *Augmented Matrix*

"Augmented matrix" is a term used in solving systems of equations. If a matrix ($m \times n$) represents the coefficients of a system of equations and a column matrix ($m \times 1$) represents the solutions of the system, the augmented matrix includes both the coefficients and the solutions in a single matrix. This form of the matrix is used in the Row Operation process of solving systems. The other two methods use the matrices separately. We will see how to do this in Section IV.

$$coefficient: \begin{bmatrix} 1 & 3 & 2 \\ -1 & 2 & 5 \\ 6 & 5 & -4 \end{bmatrix}, solutions: \begin{bmatrix} 5 \\ 24 \\ 10 \end{bmatrix}, augmented: \begin{bmatrix} 1 & 3 & 2 & 5 \\ -1 & 2 & 5 & 24 \\ 6 & 5 & -4 & 10 \end{bmatrix}$$

II. Operations on Matrices
A. Addition

To add matrices, the matrices must first be of the same size. Then, add the corresponding coordinates.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 7 & 8 \\ -3 & -2 \end{bmatrix} = \begin{bmatrix} 1+7 & 2+8 \\ 3+(-3) & 4+(-2) \end{bmatrix} = \begin{bmatrix} 8 & 10 \\ 0 & 2 \end{bmatrix}$$

B. Scalar Multiplication

Scalar multiplication is multiplying a matrix by a constant. We do this by multiplying every entry of the matrix by the scalar.

$$2\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1(2) & 2(2) \\ 3(2) & 4(2) \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix}$$

C. Matrix Multiplication

Matrix multiplication requires that the two matrices be of appropriate sizes. Unlike addition, they need not be exactly the same size. Instead the "inside" dimensions must match. For instance: an $m \times n$ matrix can be multiplied by an $n \times p$ matrix (in this order). The result will be the outside dimensions: and $m \times p$ matrix. This is so because the rows of the first matrix are multiplied by the columns of the second matrix.

$$\begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 3 & -4 \end{bmatrix} \times \begin{bmatrix} -3 & 5 & -7 & 8 \\ 1 & 6 & 4 & -5 \end{bmatrix} = \begin{bmatrix} 2(-3)+0(1) & & & \\ & & & \\ & & & \end{bmatrix}$$

The c_{11} entry of the product is the product of row 1 of the A matrix times the column 1 of the B matrix: $a_{11}*b_{11}+a_{12}*b_{21}$, here: $2(-3) + 0(1)$.

$$\begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 3 & -4 \end{bmatrix} \times \begin{bmatrix} -3 & 5 & -7 & 8 \\ 1 & 6 & 4 & -5 \end{bmatrix}$$

$$= \begin{bmatrix} 2(-3)+0(1) & 2(5)+0(6) & 2(-7)+0(4) & 2(8)+0(-5) \\ -1(-3)+1(1) & -1(5)+1(6) & -1(-7)+1(4) & -1(8)+1(-5) \\ 3(-3)+(-4)(1) & 3(5)+(-4)(6) & 3(-7)+(-4)(4) & 3(8)+(-4)(-5) \end{bmatrix}$$

$$= \begin{bmatrix} -6 & 10 & -14 & 16 \\ 4 & 1 & 11 & -13 \\ -13 & -9 & -37 & 44 \end{bmatrix}$$

Here, we multiplied a 3×2 matrix by a 2×4 matrix, resulting in a 3×4 matrix. Notice the pattern that developed in the multiplication process.

Because of this issue with the sizes of matrices, only square matrices can be switched in order and still produce a well-defined outcome; i.e. **AB** and **BA** are defined if the matrices are square; however, they need **not** be equal to each other. Only special matrices obey the commutative property, such as the zero matrix (though the size may need to change), and the identity matrix (again, the size may need to change to be well-defined), and inverse matrices (see Section E). And in most cases, even if **AB** is well-defined, **BA** will not match the correct dimensions and will not be defined at all.

The associative and distributive properties do still hold.

D. **Transpose**
The transpose of a matrix is found by exchanging the entries a_{ij} of a matrix **A** for a_{ji} of **A**^{T}. That is to say the rows will become columns and the columns rows. A matrix of the size $m \times n$ will be transposed to a matrix the size of $n \times m$.

$$\begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 3 & -4 \end{bmatrix}^T = \begin{bmatrix} 2 & -1 & 3 \\ 0 & 1 & -4 \end{bmatrix}$$

E. **Inverse**
The inverse of a square matrix is a matrix so that **A**$^{-1}$**A**=**I** and **AA**$^{-1}$=**I**. There are a number of ways to find the inverse of a matrix. Perhaps the simplest is the use of an "augmented" matrix.

$$\begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} 1 & 2 & | & 1 & 0 \\ 3 & 4 & | & 0 & 1 \end{bmatrix}$$

By building a matrix with the matrix we wish to invert on the one side and the identity matrix on the other side, we can "solve" this matrix to find the inverse matrix. The result will be a matrix with the identity on the left side and the inverse matrix on the right side. (We will see how to do this in Section IVA.) Once we find the matrix, we can check it by multiplying by the original matrix to see if it does produce the identity.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \times \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1(-2)+2\left(\frac{3}{2}\right) & 1(1)+2\left(-\frac{1}{2}\right) \\ 3(-2)+4\left(\frac{3}{2}\right) & 3(1)+4\left(-\frac{1}{2}\right) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Similarly, the check will work in the other direction as well.

For a 2×2 matrix we can also find the inverse by finding the determinant (see Section F below). This method will not work for larger matrices.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Not all square matrices are invertible, and the formula above gives a clue as to why. If the determinant of the matrix is zero, the matric cannot be inverted.

F. Determinant

Determinants are a number that is calculated from the entries in a square matrix. It can be helpful in finding inverses (of small matrices), and is useful in solving for solutions of a system of linear equations (see Section IVC below), and a number of other things. Determinants of small matrices (like 2×2) aren't too difficult, but even 3×3 matrices become quite complicated, and larger matrices are *much* more complicated. We will consider only the 2×2 case and the 3×3 case here. Larger cases, except in the case of triangular or diagonal matrices, as we will see from the 3×3 case, can be reduced to the smaller case. In triangular or diagonal matrices, only the multiplication along the main diagonal survives, so in this special cases, large matric determinants can also be found easily.

In the 2×2 case, the formula for the determinant is det(A)=$a_{11}*a_{22}-a_{12}*a_{21}$. Another way to think about this formula is to multiply along the main diagonal, and subtract the product along the second diagonal.

$$\det\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\right) = 1(4) - 2(3) = 4 - 6 = -2$$

In the case of a 3×3 matrix, the formula has nine matrix entries in it, and six terms. Instead of writing it out in terms of the matrix entries, we can write out in terms of the first row and "sub-matrices". Consider the matrix below:

$$\det\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right) = a_{11}\det\left(\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}\right) - a_{12}\det\left(\begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}\right) + a_{13}\det\left(\begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}\right)$$

The first term can be obtained by the row 1-column 1 term times the value of the determinant for the matrix obtained by ignoring the first row and the first column. Similarly, the second term is obtained by the row 1-column 2 term and the determinant of the entries left when the first row and the second column are covered up. The third term is likewise obtained from the row 1-column 3 term and the determinant obtained from covering up the first row and the third column.

Filling in the entries of the matrix with values we get:

$$\det\left(\begin{bmatrix} 1 & 4 & -2 \\ 2 & 9 & 3 \\ -3 & -5 & -1 \end{bmatrix}\right) = (1)\det\left(\begin{bmatrix} 9 & 3 \\ -5 & -1 \end{bmatrix}\right) - 4\det\left(\begin{bmatrix} 2 & 3 \\ -3 & -1 \end{bmatrix}\right) + (-2)\det\left(\begin{bmatrix} 2 & 9 \\ -3 & -5 \end{bmatrix}\right)$$

$$= (1)\left[9(-1)-(3)(-5)\right] - 4\left[2(-1)-(3)(-3)\right] - 2\left[2(-5)-9(-3)\right] = -24-28-34 = -86$$

Some properties of determinants:
- Switching a row or column with another in the matrix will change the sign of the determinant.
- If the matrix determinant is zero, then the matrix is not invertible.
- A strategy for finding the determinant of very large matrices is to use row operations (described in Section G below) to convert the matrix to an equivalent triangular one. This may be easier than working out the formula using sub-matrices.
- Notice that the terms of the 3×3 determinant are +/-/+. In the 4×4 case would have signs of +/-/+/-, and so on for larger matrices.
- In the formula above, we used row 1 as the basis for the formula, however, it works just as well using column 1 or any other row or columns in the matrix.
- Multiplying a row by a scalar multiplies the determinant by that scalar. Multiplying a matrix by a scalar is like multiplying the determinant by c^n (where n is the size of the matrix).

G. Row Operations

Row operations are a process that we will use below to solve a system of equations represented as a matrix. There is a similar (and related) process of column operations which we will not cover explicitly.

Row operations is a process related to elimination by addition in a system of equations. There are three possible row operations.
- Switch two rows in a matrix.
- Multiply a row by a scalar.
- Add two rows together (with or without multiplying by a scalar).

Consider the matrix we had that was formed to help us find the inverse of a square matrix.

$$[A \quad I] = \begin{bmatrix} 1 & 2 & | & 1 & 0 \\ 3 & 4 & | & 0 & 1 \end{bmatrix}$$

To use row operations to solve this matrix, we'd like to get the right side to look like the identity matrix, and then the right side will be the inverse. We do this by row operations.

For the identity matrix, we want the first entry (a_{11}) to be 1, so that much is okay (if it were not 1, we could divide all the entries by that number). We'd like to next eliminate the 3 in the second row. To achieve this, I will add (-3)(row 1) to the second row = [-3 -6 -3   0] +[3 4 0 1] = [0 -2 -3 1]. The matrix now looks like this:

$$[A \ \ I] = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{bmatrix}$$

$$-3R_1 + R_2 \Rightarrow R_2 \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{bmatrix}$$

To clear the 2 in top row, second column, I can add these two rows together:

$$R_1 + R_2 \Rightarrow R_1 \begin{bmatrix} 1 & 0 & -2 & 1 \\ 0 & -2 & -3 & 1 \end{bmatrix}$$

Lastly, divide row 2 by -2.

$$-2R_2 \Rightarrow R_2 \begin{bmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & \dfrac{3}{2} & -\dfrac{1}{2} \end{bmatrix}$$

Now, the last two columns of the matrix are the inverse of A that we had used in our example.

These operations highlighted two of the row operations. The last one, switching rows, is generally used in cases like below:

$$\begin{bmatrix} 3 & -2 & -1 \\ 0 & 4 & -3 \\ 1 & 5 & 2 \end{bmatrix}$$

In most cases, we want the entry in the first row and first column to be 1. Here, we could choose to divide by 3, and then work with fractions, or else, move the third row up to be row 1. When doing these calculations by hand, this is easier for humans. Your calculator will choose the brute force approach. If the entry is 0, even the calculator will swap rows.

III. **Converting Systems of Equations to Matrix Equations**
In order to solve a system of equations by matrix methods, we first need to convert the system to matrices. Depending on the method we are using we will either need one augmented matrix (row operations), or three matrices (inverses or Cramer's Rule).

Given a system of equations, as shown below, the augmented matrix contains the coefficients of the variables, and the solutions to the equations. They are typically divided by a line.

$$\begin{cases} 3x - y + z = 6 \\ 4x + y - z = 8 \\ -2x - 3y + z = 10 \end{cases} \Leftrightarrow \begin{bmatrix} 3 & -1 & 1 & 6 \\ 4 & 1 & -1 & 8 \\ -2 & -3 & 1 & 10 \end{bmatrix}$$

To write the same system of equations in three matrices, we have one matrix for the coefficients of the variable, one column matrix for the variables, and one column matrix for the solutions. We can test by matrix multiplication to see that this representation results in the original system.

$$\begin{cases} 3x - y + z = 6 \\ 4x + y - z = 8 \\ -2x - 3y + z = 10 \end{cases} \Leftrightarrow \begin{bmatrix} 3 & -1 & 1 \\ 4 & 1 & -1 \\ -2 & -3 & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \\ 10 \end{bmatrix}$$

For the last set up, the coefficient matrix is usually labeled **A**, the variable matrix **X**, and the solution matrix **B**.

IV. **Solving Systems of Equations by Matrix Operations**
  A. **Solving by Row Operations**
  To solve a system by row operations, like with the identity matrix, we want the coefficient portion of the matrix to be the identity matrix, or at least triangular. Which one depends on how hard we want to work.

$$\begin{cases} 3x - y + z = 6 \\ 4x + y - z = 8 \\ -2x - 3y + z = 10 \end{cases} \Leftrightarrow \begin{bmatrix} 3 & -1 & 1 & 6 \\ 4 & 1 & -1 & 8 \\ -2 & -3 & 1 & 10 \end{bmatrix}$$

$$R_2 - R_1 = R_1 \Rightarrow \begin{bmatrix} 1 & 2 & -2 & 2 \\ 4 & 1 & -1 & 8 \\ -2 & -3 & 1 & 10 \end{bmatrix}$$

$$\begin{matrix} R_2 - 4R_1 = R_2 \\ R_3 + 2R_1 = R_3 \end{matrix} \Rightarrow \begin{bmatrix} 1 & 2 & -2 & 2 \\ 0 & -7 & 7 & 0 \\ 0 & 1 & -3 & 14 \end{bmatrix}$$

$$\begin{matrix} R_3 = R_2 \\ \left(\dfrac{1}{7}\right) R_2 \end{matrix} \Rightarrow \begin{bmatrix} 1 & 2 & -2 & 2 \\ 0 & 1 & -3 & 14 \\ 0 & -1 & 1 & 0 \end{bmatrix}$$

$$R_2 + R_3 = R_3 \Rightarrow \begin{bmatrix} 1 & 2 & -2 & 2 \\ 0 & 1 & -3 & 14 \\ 0 & 0 & -2 & 14 \end{bmatrix}$$

$$\left(-\dfrac{1}{2}\right) R_3 \Rightarrow \begin{bmatrix} 1 & 2 & -2 & 2 \\ 0 & 1 & -3 & 14 \\ 0 & 0 & 1 & -7 \end{bmatrix}$$

We've reached a triangular (coefficient) matrix. This is sufficient to solve the system by hand by back solving. The bottom row says that $z = -7$, the row above it $y - 3z = 14$, and we can find y by substitution: $y = -7$. And solving for x from the top row $x + 2y - 2z = 2$, we get $x = 2$. The form of this matrix is also called row echelon form. This can be applied to matrices of any size where the diagonal is all ones, and the entries below the diagonal are all zeros, even if the matrix is not square.

If we don't wish do back solve this way, we can continue until we've reached a matrix with an identity matrix in place of all the coefficients. This form is called row-reduced echelon form. From this form, we can read the values of the variables directly off the matrix.

$$\begin{bmatrix} 1 & 0 & 0 & | & 2 \\ 0 & 1 & 0 & | & -7 \\ 0 & 0 & 1 & | & -7 \end{bmatrix}$$

Which we use depends on whether we think that back-solving is easier than row-reducing. But, if we are letting a computer do the work for us, the row-reduced form is better since the computer will be more reliable in its arithmetic than we will be (and faster).

B. Solving using Inverses
To solve a system using matrices, we use the analogy of a matrix equation **AX=B**. If this was a simple linear equation, I'd divide the equation by the coefficient of the variable. There are no rules for "dividing" matrices, but instead, we can multiply by the inverse (just as division is really multiplication by the reciprocal or the multiplicative inverse): **X=A⁻¹B**.

This is not an efficient process to do by hand (we have to row reduce to find the inverse matrix (except in the case of the 2×2 matrix for which we have a formula), and then we have to do multiplication by hand). However, for a computer, this process is quite simple.

This form of the solution is typically used in proving theorems about a system.

C. Solving by Cramer's Rule
Cramer's Rule is a method that allows us to solve for one variable at a time. The procedure involves the determinant of the coefficient matrix, and the determinant of a matrix where the column representing the variable we wish to solve for has been replace by the column of the solution matrix. Let's call this matrix **K**$_i$.

Let's solve for the y-variable in the matrix we solved by row reducing. We need the determinant of **A**.

$$\det\left(\begin{bmatrix} 3 & -1 & 1 \\ 4 & 1 & -1 \\ -2 & -3 & 1 \end{bmatrix} \right) = -14$$

We also need the determinant of **K**.

$$\det\left(\begin{bmatrix} 3 & 6 & 1 \\ 4 & 8 & -1 \\ -2 & 10 & 1 \end{bmatrix}\right) = 98$$

This **K** is the **A** matrix but with the 2nd column (representing the y-variable) with the solution values. Then we find that $y = \dfrac{\det(K)}{\det(A)} = \dfrac{98}{-14} = -7$. Following this procedure, we can find the values of the other variables as well, but to solve for x, the first column is replaced by the solution matrix, or to solve for z, the third column is so replaced.

In the next lecture, we'll build simple linear regression models, including putting some of our linear algebra to use.

References:
1. https://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf
2. https://assets.openstax.org/oscms-prodcms/media/documents/IntroductoryStatistics-OP_i6tAI7e.pdf
3. https://faculty.ksu.edu.sa/sites/default/files/probability_and_statistics_for_engineering_and_the_sciences.pdf
4. https://cran.r-project.org/web/packages/faux/vignettes/rnorm_multi.html
5. https://stackoverflow.com/questions/53972889/how-do-i-simulate-two-variables-in-r-which-are-non-linearly-related
6. https://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf
7. https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/spearman-rank-correlation-definition-calculate/
8. https://www.statisticshowto.com/kendalls-tau/
9. http://betsymccall.net/prof/courses/mathnotes/handouts/algebra/130matrices.pdf