Foundations of Data Analysis

The University of Texas at Austin

R Tutorials: Week 1

Indexing

In this R tutorial, we're going to learn about indexing and using logical statements to pull out certain data from a vector, based on some criteria that we set. So we use parentheses when we want to use a function and have some arguments. We're going to use the square brackets to index a vector. So if I want to pull out the second value in the vector, "myvariable", I'm going to use square brackets and ask for 2.

```
myvariable[2]
```

```
## [1] 28
```

This is basically telling R give me the second number in that vector, right? And I see that 28 is the second number in the object, "myvariable". So beyond just asking for a specific index, we can actually ask for the cases from "myvariable" that meet some specific criteria. If I give R a logical statement, such as "myvariable" is greater than 0, now I know that I have four positive numbers and one negative number in the object my variable. If I give it this logical statement, right "myvariable" greater than zero, I'm going to see a true/false output here.

```
myvariable > 0
```

```
## [1]  TRUE  TRUE FALSE  TRUE  TRUE
```

It's basically looking at this criteria, and then going into the vector, "myvariable", and saying, is this statement true for each number? We see that it's true for the first number, which is positive 72, right? 72 is greater than 0, so is 28. However, negative 9 is false. It's not greater than zero, and then we get also a true and a true for the last two positive numbers. If we include this logical statement in an indexing, so if I ask for "myvariable" where "myvariable" is greater than 0, instead of just pulling out, say, the second number of that vector, it's going to pull out wherever this is true, and give me those numbers in return.

```
myvariable[myvariable > 0]
```

```
## [1] 72 28 12 11
```

So what this is going to give me are the four positive numbers from my vector. The other types of logical statements I can use are the symbols less than, greater than, equal to, which is 2 equal signs back to back, not equal to, which is an exclamation point and an equal sign.

Figure 1: R logicals.

So just to see what I'm talking about here, if I want to take the cases of "myvariable", where "myvariable" is equal to 12, I'm going to include this double equal sign. That's telling R it's a logic statement. Run this and I'm going to get the single case where my vector equals 12.

```
myvariable[myvariable == 12]
```

```
## [1] 12
```

If I ask for the same thing, but instead of equal to I say not equal to, I'm going to get the other four numbers from my vector.

```
myvariable[myvariable != 12]
```

```
## [1] 72 28 -9 11
```