Baseball in Thin Air: Using Pivot Tables to Analyze the 2015 Colorado Rockies

In American baseball, teams are often known for the unique characteristics of their stadiums. In Boston, the Red Sox play in Fenway Park, which is known for a giant outfield wall known as “The Green Monster”. In Chicago, the Cubs play in Wrigley Field, known for walls covered in ivy.

The stadium in Denver, however, is known not so much for its physical characteristics but the atmosphere. The elevation of Coors Field is 5,200 feet above sea level. The air is so thin, the ball flies farther. And that in turn leads to more runs.

Unfortunately for the Rockies, the air is thin for BOTH teams. In this exercise, we’ll explore just how much the air impacts the team. We’ll look at a log of games played by the team in 2015, and calculate how the team performs in their home stadium vs games played on the road.

**Importing the Data**

Let’s take a look. The game log data is on the Web. You can open files on the Web directly in Excel. Go to File/Open URL and enter this address:

<http://bit.ly/1NQ1pSV>

The text import wizard will appear. In step 1, It guesses that we’re trying to open a text file where the columns are delimited by something ( commas), which is correct, so we can click “next”.

In step 2, we confirm that our delimiter is in fact a comma. In many cases, data will also come wrapped in “” marks. This is not true in this case, but it’s important to be aware of this possibility.

Step 3 is where we can make sure our data is optimized for analysis. One of the big mistakes you can make here involves data types. Excel tries to guess whether each column is a number, date or text but it’s smart to review the choices.

Scanning the data we’re about to import, you can see one possible problem, the column labeled “W-L” – that’s the won-loss record of the team after every game. The dash is a problem – Excel will likely interpret that as a minus sign or a date signifier, so we need to specify during our import that this column is text.

Once all of the column types look ok, we can go ahead and finish the wizard and our data is now in the spreadsheet. Keeping the good practice of saving this as the original data and making a working copy, hit cmnd+A to select everything, cmnd+c to copy it, make a new worksheet, and hit cmnd+v to paste it in. Now save the file to your computer.

**Cleaning the Data**

So now we have our import, let’s think of some questions we might want to ask as we interview the

|  |
| --- |
| TODO: write exercise, do thesee calcs? |
| 1) create new w/l column left(1) |
| 2) )Home-road pivot based on column f -- need header? Change symbol? |
| 3) runs scored against by home road; average per game? |
| 4) Record by month? By day of week? By day/night? |
| 5) average, min, max attendance by home/road, day of week (doubleheader problem) |
| 6) longest winning streak? Losing streak? |