Chapter 2

# Try It Yourself: Clean Incorrect Formats

## Excel Instructions

1. Open the vote.csv data set in Excel.
2. Verify that the favorite\_pet variable is in Column E.
3. Calculate totals before cleaning the data.
   1. Use the COUNTIF function in an empty cell to determine how many people indicated that their favorite pet was a “dog”:

=COUNTIF(E:E,"dog")

* 1. Use the COUNTIF function in an empty cell to determine how many people indicated that their favorite pet was a “cat”:

=COUNTIF(E:E,"cat")

* 1. Record these counts.

1. Use FIND AND REPLACE to clean the data.
   1. From the home tab, choose Find & Select → Replace.
   2. In the ‘Find what’ field, type “dogs”, and in the ‘Replace with’ field, type “dog”.
   3. Click ‘Replace All’
   4. Repeat this process, but in the ‘Find what’ field, type “cats”, and in the ‘Replace with’ field, type “cat”.
2. Use an IF statement for further cleaning.
   1. Right-click on Column F and choose Insert.
   2. Title the new blank Column F “favorite\_pet\_cleaned”.
   3. Type into cell F2, the following formula to replace words that start with D (d-) and end with G (-g) to “dog” and words that start with C (c-) and end with T (-t) to “cat”: =IF(AND(LEFT(E2,1)="d",RIGHT(E2,1)="g"),"dog",IF(AND(LEFT(E2,1)="c",RIGHT(E2,1)="t"),"cat",E2))
   4. Apply this formula to all rows for this column.
3. Calculate totals after cleaning the data.
   1. Use the COUNTIF function in an empty cell to determine how many people indicated that their favorite pet was a “dog” in the cleaned data: =COUNTIF(F:F,"dog")
   2. Use the COUNTIF function in an empty cell to determine how many people indicated that their favorite pet was a “cat” in the cleaned data: =COUNTIF(F:F,"cat")