Chapter 2

# Try It Yourself: Clean Incorrect Formats

## StatCrunch Instructions

1. Open the vote data in StatCrunch
2. Stat → Tables → Frequency
3. Select Columns → favorite\_pet
4. Click Compute!
5. Sort the results by descending frequency to see the most frequent.

Clean the 'Favorite\_Pet' column for simple replacements:

1. Edit → Find and Replace
2. In “Find” type **dogs**
3. In “Replace with” type **dog**
4. Find in → Selected Columns → Select favorite\_pet
5. Replace all
6. In “Find” type **cats**
7. In “Replace with” type **cat**
8. Find in → Selected Columns → Select favorite\_pet
9. Replace all

Advanced cleaning for conditional replacements:

1. Data → Compute → Expression
2. Enter the following expression:

*ifelse(startsWith(Favorite\_pet,d) and endsWith(Favorite\_pet,g),dog,ifelse(startsWith(Favorite\_pet,c) and endsWith(Favorite\_pet,t),cat,Favorite\_pet))*

1. Name the new column cleaned\_favorite\_pet
2. Click Compute!

Count occurrences post-cleaning:

1. Stat → Tables → Frequency
2. Select Columns → cleaned\_favorite\_pet
3. Click Compute!