Recreating VLOOKUP in R

Nov 13, 2019 · R

I recently gave a talk on Excel that focused on pivot tables and the VLOOKUP function. Recreating the power of VLOOKUP is easy in R using the dplyr package from the tidyverse.

```
# Load dplyr package
library(dplyr)
```

One of the examples I used was collapsing research ratings into more broad "buckets" for principal, major, and leadership gift-rated prospects. We'll start be making a rating lookup table that contains all our ratings along with a corresponding "bucket." This could also be stored in a .csv or .xlsx file and loaded in separately.

```
rating_lookup <-
  as_tibble(list(
    rating = c(
      "A - $2.5M to $5M",
      "B - $1M to $2.5M",
      "C - $500,000 to $1M",
      "D - $250,000 to $499,999",
      "E - $100,000 to $249,999",
      "G - $50,000 to $99,999",
      "H - $25,000 to $49,999",
      "I - $10,000 to $24,999"
    ),
    rating_bucket = c("PG", "PG", "MG", "MG", "MG", "LG", "LG", "LG")
  ))

# view our new lookup table
rating_lookup
```

```
## # A tibble: 8 x 2
##   rating                   rating_bucket
##   <chr>                    <chr>
## 1 A - $2.5M to $5M         PG
## 2 B - $1M to $2.5M         PG
## 3 C - $500,000 to $1M      MG
## 4 D - $250,000 to $499,999 MG
## 5 E - $100,000 to $249,999 MG
## 6 G - $50,000 to $99,999   LG
## 7 H - $25,000 to $49,999   LG
## 8 I - $10,000 to $24,999   LG
```

Now, we'll make a sample dataset with ID numbers along with ratings.

```
df <-
  as_tibble(list(
    # create an id column using the numbers 1 through 50
    lookup_id = 1:50,
    # create a rating column by randomly sampling our previously generated ratings 50 times
    rating = sample(rating_lookup$rating, size = 50, replace = TRUE)
  ))

# look at only the first 10 observations
head(df, n = 10)
```

```
## # A tibble: 10 x 2
##    lookup_id rating
##        <int> <chr>
##  1         1 H - $25,000 to $49,999
##  2         2 B - $1M to $2.5M
##  3         3 C - $500,000 to $1M
##  4         4 E - $100,000 to $249,999
##  5         5 I - $10,000 to $24,999
##  6         6 A - $2.5M to $5M
##  7         7 H - $25,000 to $49,999
##  8         8 D - $250,000 to $499,999
##  9         9 B - $1M to $2.5M
## 10        10 D - $250,000 to $499,999
```

Now that we have our dataset as well as our lookup table, we can merge our rating buckets into our dataset using the `left_join` function.

```
df <- left_join(df, rating_lookup, by = "rating")

# look at only the first 10 observations
head(df, n = 10)
```

```
## # A tibble: 10 x 3
##    lookup_id rating                   rating_bucket
##        <int> <chr>                    <chr>
##  1         1 H - $25,000 to $49,999   LG
##  2         2 B - $1M to $2.5M         PG
##  3         3 C - $500,000 to $1M      MG
##  4         4 E - $100,000 to $249,999 MG
##  5         5 I - $10,000 to $24,999   LG
##  6         6 A - $2.5M to $5M         PG
##  7         7 H - $25,000 to $49,999   LG
##  8         8 D - $250,000 to $499,999 MG
##  9         9 B - $1M to $2.5M         PG
## 10        10 D - $250,000 to $499,999 MG
```

Finally, we can use dplyr's `count` function to recreate some of the power of Excel's pivot tables.

```
count(df, rating_bucket, name = "prospects")
```

```
## # A tibble: 3 x 2
##   rating_bucket prospects
##   <chr>             <int>
## 1 LG                   21
## 2 MG                   16
## 3 PG                   13
```

excel dplyr