

From: Bowman, Randal
To: [Hoy, Mark](#)
Cc: [Shulman, Stu](#); [Marcia Cash](#)
Subject: Re: problem in putting dups into buckets & datasets
Date: Tuesday, August 01, 2017 9:07:00 PM
Attachments: [image.png](#)

Perhaps I didn't explain clearly, my response in all caps after your questions -

if I gather correctly - you are trying to create a bucket of all the exact dups (only 1 copy from each group) and the singles, correct? NO, I AM TRYING TO CREATE A BUCKET AND THEN DATASET (NOT INITIALLY REALIZING I COULD GO STRAIGHT TO DATASET W/O A BUCKET) OF ALL OF THE COMMENTS IN EACH OF CHECKED EXACT DUPLICATE GROUP [I.E. ALL COMMENTS IN GROUPS 1-10], NO SINGLES AT ALL, AND WAS INSTEAD GETTING ONLY ONE COPY FROM EACH GROUP. THE PURPOSE IS TO GET THEM INTO "OPPOSE" "SUPPORT" AND "UNCODABLE" GROUPS SO WE CAN COUNT THEM, NOT TO ACTUALLY CODE THEM. WHEN I SAID "CODE" IN EARLIER MESSAGE I MEANT "COUNT"

Are you using the check box and trying to create a bucket using the "Add to new bucket" -> "Checked Items" (as shown below) for this? YES. I LEFT AN EXAMPLE OF EACH IN THE BUCKETS AND DATASET GROUPINGS AS EXPLAINED ABOVE.

FOR TOMORROW, I WILL WANT TO CREATE A DATASET WITH ALL OF THE SINGLES, AND PRESUME "ALL SINGLE ITEMS" WILL DO THAT.

On Tue, Aug 1, 2017 at 9:02 PM, Hoy, Mark <mark@texifter.com> wrote:

Hi Randal -

I just got home from my drive and am looking at this.

I think I understand what you are trying to do, so, let me take a stab to see if we're on the same page... if I gather correctly - you are trying to create a bucket of all the exact dups (only 1 copy from each group) and the singles, correct? Are you using the check box and trying to create a bucket using the "Add to new bucket" -> "Checked Items" (as shown below) for this? If so - this will only add those items that are checked to the bucket. If you want all the singles, use the "all items" option instead of the "checked items" - and I believe this should get you what you need.

(and please pardon me if I'm not correctly understanding the issue) - hope this helps.

- Mark

Inline image 1



On Tue, Aug 1, 2017 at 7:06 PM, Shulman, Stu <stu@texifter.com> wrote:

Randy,

You caught us on a bad day. Mark is returning from a (b) (6) and I am dealing with an (b) (6). We have not forgotten you, it has just not been possible for me to be at my desk today as I am trying to (b) (6)

I can definitely get you coding my tomorrow, (b) (6)

Stu

On Tue, Aug 1, 2017 at 5:43 PM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

now it is happening with the near-duplicates. I've tried twice to create a cluster 11-50, less cluster 25, database, and both time the results had 39 items, not the contents of the clusters. I left the last database from this attempt as well, so the first 2 databases under July 20 are failures, the first being a test with one item checked after several failures, and the 3rd is the successful near-duplicate result. All were done in the same way, as set out in the previous email.

We are running into time-to-completion concerns, and need to be at least able to start coding on the singles tomorrow morning.

On Tue, Aug 1, 2017 at 2:36 PM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

I have a sense this happened before, but neither Marsha nor I can resolve it.

I started preparing to code the July 20 dups and clusters. However, when I check some dup groups and use the "settings" symbol to move the checked duplicate groups to a bucket, only the first item of that set is captured - i.e. if I check items 1-8 and create a bucket, that bucket has 8 items in it, not the thousands contained in the checked groups. If you look at the July 20 buckets you will see 2 examples that I didn't delete.

After this happened the first time I tried checking "include all duplicates" as part of the process for creating the bucket, but ended up with a bucket that had 273,000 items in it, so that is clearly not the solution. I then tried going straight to datasets, without buckets, but same results - see datasets.

For some reason, this does not seem to be a problem with the near-duplicates, as I just tried those instead and the dataset I created using the exact same process has the numbers that it should. I thought earlier that I had tried both dups and nears earlier, and advised Marsha that way, but apparently I had only tried dups. There is one nd dataset with 12,000 items as an example of a successful creation.

Additionally, I thought I should check on how to put the 55,920 singles into a dataset

so coding those could start. Do I using the "settings" to "add to new dataset" "all single groups"? And should I go bucket then dataset, or is it OK to go straight to datasets?

--

Dr. Stuart W. Shulman

Founder and CEO, Texifter

Cell: [413-992-8513](tel:413-992-8513)

LinkedIn: <http://www.linkedin.com/in/stuartwshulman>

⚙️ Items per page: 20 ▼ Page: 1 2 3 4 5 6 7 8

Add to new bucket »
Add to existing bucket »
Add to new dataset »

Checked items
All items
All single groups

Delete groups

asured national monuments
DOI-2018-05 00853

that our treasured national monuments are up for review