

Digital Maestro
m a g a z i n e

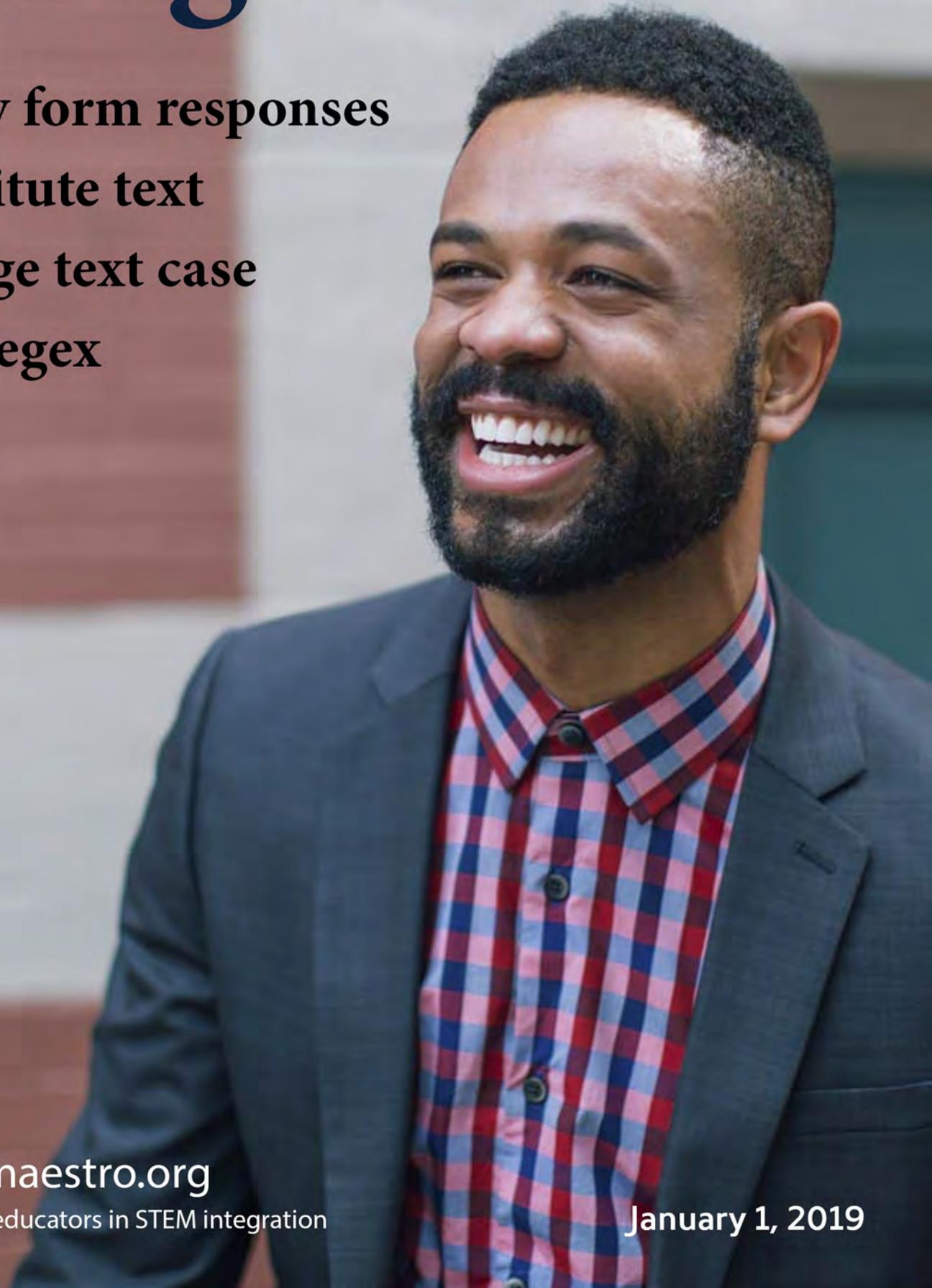
Google Sheets

Query form responses

Substitute text

Change text case

Use Regex



digitalmaestro.org

Supporting educators in STEM integration

January 1, 2019

Table of Contents

Introduction	4
Google Forms Data	6
Cleaning data	14
Review	25

Introduction

This is the second part of my five-part series. The first part, released on December 15, 2018, covered the collection of data with Google Forms. The form did more than simply collect information. It verified and validated the information provided by presenters, participants, and vendors.

In this issue, we will look at the collected information. We took steps to validate the information on the form. The validation didn't include spell checking or formatting. Some of the formatting issues we encounter include capitalization, unwanted spaces, and misspellings. This information needs to be reviewed before it appears in the presentation schedule. This corrected information is sent to the presenters for verification.

In these lessons, we will import the form data into another sheet. I don't like to work with the collected information directly. The imported information will be manipulated so we have consistency. This includes spell checking and letter case verification. The data will form the basis for the schedule. The schedule will import the information from our verification sheet. Every step of the import process keeps a consistent link to the original information.

In the beginning correct the original information. I quickly learned this was a mistake.

The information shown in these lessons is not actual information from a scheduled event. We need to look at email addresses and campus information. To keep things anonymous I have created a Google Sheet that resembles collected information. I will provide a link to the data in this lesson.

Skills

Here are some of the skills you will learn in these lessons. The first step is to pull the information we want to clean. This information will be used to construct the schedule. The information is pulled into another sheet using a function called QUERY. A query requests data from a database table. The sheet in our Google Sheet is a table with data.

The query will request data from select columns in the table. Each column in the table is referred to as a field. The terms query and field come from the world of databases. Databases include one or more tables. We refer to all the information in each row as a record. Records are part of databases.

Cleaning data is a common part of collecting data. Scientists and researchers do this all the time. I use several functions to clean the collected data. Functions are special instructions used to carry out a series of steps.

I use the SUBSTITUTE function to replace unwanted characters. These characters often include periods, dashes, or extra letters. The substitute function looks for content I choose and replaces it with content I also choose. For example, the substitute function will learn for a character like a comma. It will replace the comma with another character of my choosing, like a period.

The TRIM function removes unwanted characters. It does this much like the SUBSTITUTE function. I use the function to trim extra spaces. It doesn't remove all spaces. It is smart enough to know that two spaces in a row are not common. The extra space is removed. I like the function because it removes spaces from the beginning and end of words. Extra spaces cause formatting issues on published content.

Nested functions are one way to use more than one function at a time. I use the trim function within the substitute function. This removes unwanted spaces and removes unwanted characters.

I am a fan of something called Regular Expressions. Regular expressions use patterns to search for information. Searches are combined with REPLACE or EXTRACT. Replace with Regex, Regular Expressions works much like the substitute function. I prefer Regex whenever possible because it affords better precision.

The substitute function is literal. Here is an example. We want to add a space after a period if the space character is missing before the next sentence. Substitute will search for a period and then the letter "A" for example. What if the period is followed by another letter? This is where Regex really shines. A regular expression can be formatted to search for a period and any letter or number.

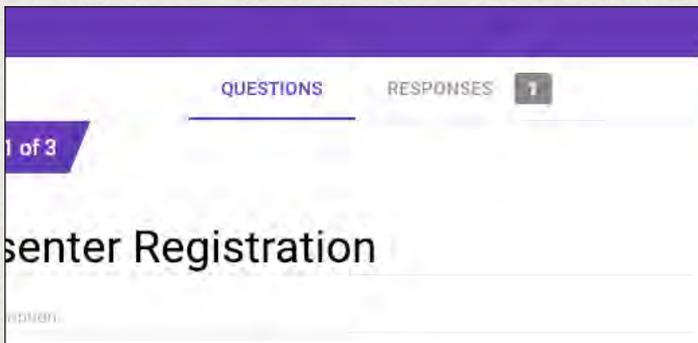
The PROPER function is used to reformat text with the proper case. The first letter in a word is changed to uppercase. The rest of the letters are changed to lowercase. That is all we will be using it for in our lesson. This simple function saves me lots of time.

Cell references are a staple of working with spreadsheets. The functions we use to take advantage of cell references. Cell references fall into two categories. Cell references are either relative or absolute. Relative cell references change with reference to the function. Absolute cell references remain fixed to one or more cells. We will use both forms in our lessons.

Google Forms Data

Google Forms collects all the responses and arranges them within the form. This data is seen in the responses section of the form. We need to get this data out of the Forms environment and to a spreadsheet. Open the Form with the presenter responses. Go into edit mode if you are not already in that mode. Edit mode allows you to update the form.

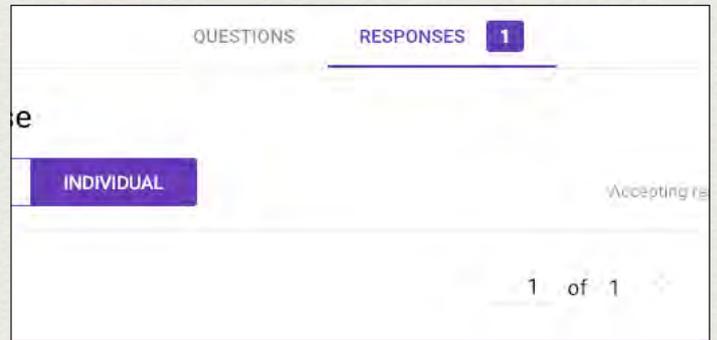
My form doesn't have any responses other than my own for testing. I created a sheet with several records of fake presenter responses. The names and emails in the form are not representative of anyone in my district.



Click the Responses tab in the form. In the responses section, we have several graphs representing the submitted information. The information collected is displayed with nice graphs. The information displayed this way helps to visualize the presentation submission.

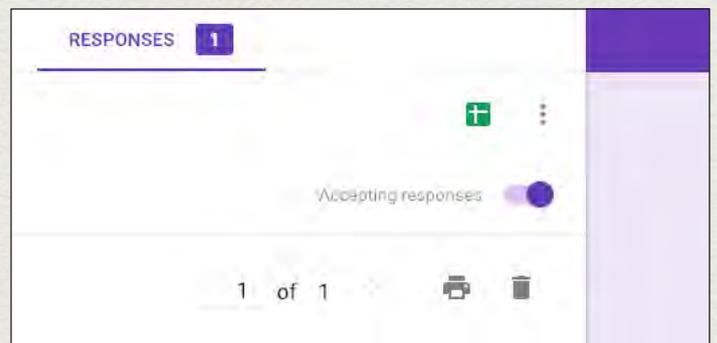
The information is helpful after the conference. It is a good way to see which presentations were attended most.

Session attendance is gathered from scanned badges at each session.



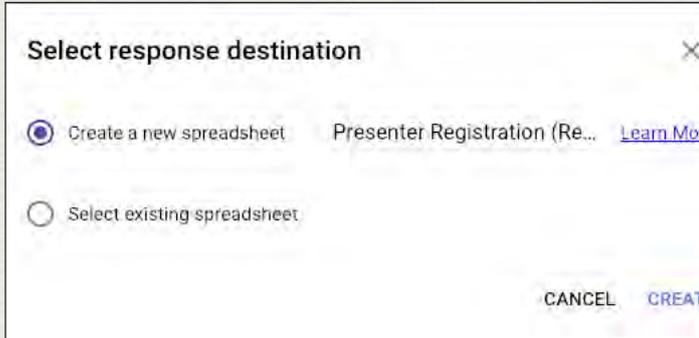
When the conference is over I like to switch off the accepting responses option. This prevents anyone from adding a presentation without us knowing about it. We do get people that want to submit their proposal at the last minute. We rarely turn anyone away. We do like to keep tabs on anyone that submits proposals after the due date.

Above the accepting responses button is a green square icon. This icon represents a spreadsheet. Click the spreadsheet icon.



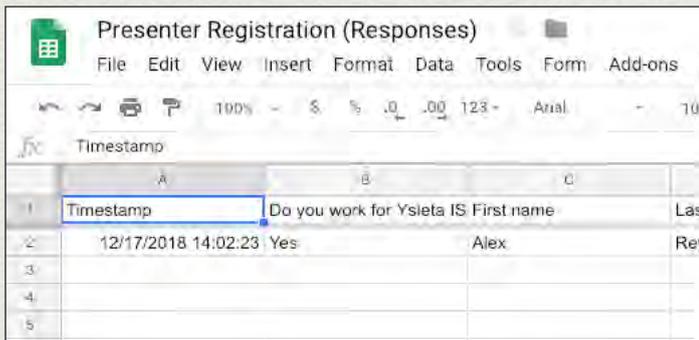
We are presented with a response destination selector. Google Forms will create a spreadsheet with the collected

responses. The name of the sheet is the same name as the form. The Sheet is titled Presenter Registration (Responses). Click the Create link.



The spreadsheet is created and opened in a separate tab. The spreadsheet contains the titles of all the questions in the form. The titles are displayed in the order they appear in the form. The first field in the sheet is a timestamp. This is not part of our questions. It is included by Google Forms. The timestamp marks the date and time when the form was submitted.

The timestamp information is often used for early bird discounts. Campuses that have teachers register early receive a discount on the conference price. The discount is given to participants. This is another reason for scanning badges. Presenters do not pay for the attendance.



The form is empty so I will fill the form in with 50 presenters. This is the average number of presenters we have at the conference.

Irregular form data

The form has information that looks odd when printed and displayed on the schedule. One of these common oddities is the use of all caps for their name.

	LEON	BASS	LBass@yisd.net

Sometimes they don't bother to use capital letters at all.

	melba	stone	MStone@yisd.net

Other oddities are not obvious. Spaces at the beginning of names are one of these not so obvious oddities.



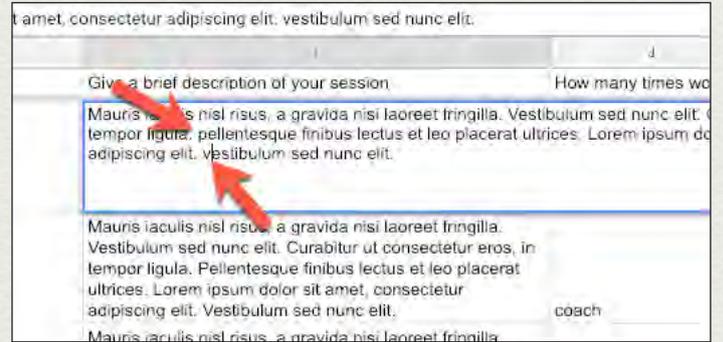
Random characters typed as part of regular information. There is a rogue dot after Wilkin's last name.



Presentation descriptions often have misspelled words. They are also lacking in good punctuation. This one is missing a capital letter in several places. We read all presentation submissions. I use Grammarly and HemingwayApp to cleanup presenter titles and descriptions. I try to keep the essence of the description wherever possible.

No you aren't having trouble reading the text. I have used Lorem Ipsum text. This is filler text used to represent finished text in publications. It is commonly used in magazines and newspapers to help

design the final product. The filler text is eventually replaced with the actual text.



We controlled the format of a lot of the form information. There isn't much for us to fix. That was the whole purpose of the design and validation of information in the form.

Querying form data

We won't correct the information in the form directly. We will correct the information from the queried information. I don't like to touch the original information. I like to keep it as a reference. I need to take a look back at the original information if a question arises. This helps to clarify questions regarding a presenters title, description, or contact information.

We will query the information we need. Querying information pulls the information we want from the original data. Queries are very useful tools. The query will extract just the information we need.

The information we need for this query is related to the scheduling of sessions. For the schedule, we need the presenter's

Digital Maestro Magazine

Thank you for reading these sample pages.

Purchase the full issue at...

digitalmaestro.org/shop

Digital Maestro Magazine