Youth Finance Institute of America

Data Analysis Workflow

A review on the updated data collection and analysis process for YFIA in-person workshop insights

Ishika Prashar, Summer 2023 Intern

Background/Context

In person workshops are an integral part of YFIA, and data is collected from attendees in various stages through physical forms which include:

1. Pre-test & Post-test
    - Evaluate prior knowledge (pre-test) and improvement after workshop (post-test)
2. Survey
    - Measure workshop positivity (rating ≥ 4)
    - Measure perceived learning
    - Measure confidence in teaching topic to others
    - Capture attendee age

Other Demographics are collected through visual observation which captures perceived racial representation and gender distribution.

Data Usage

Prior Process

There is a spreadsheet used to record results, but it is manual input and summarizes most of the data rather than comprehensively include all features. Calculations such as averages and survey ratings are done by hand.

Issues of Concern

- Manual calculations can be flawed/inaccurate
- Inefficient
- Issues with data storage
  - For example, if the physical forms are lost or damaged, there is no way to double check results or attendance at a later time

Improved Process

In response to identified concerns, and with the aim of enhancing the data collection and assessment process, we have introduced a fresh spreadsheet template. This template includes comprehensive instructions, designed to cater to individuals with varying levels of familiarity with Google Spreadsheets and allow any YFIA member to collect and analyze data.

There are three tabs/sheets inside the main spreadsheet:	
1	Workshop Data
2	MM/DD/YY Workshop
3	Summary

With this improved format, data can be stored long term, calculations are automated, and more insights are extracted from the data itself.

Workshop Data Sheet

- Data repository where all workshop data is manually entered
- Separations are created between Name/Age and test/survey results columns to maintain impartiality and ensure an objective analysis (Name/Age has no correlation with test and survey data of corresponding row)
- The "date" column requires the use of the date function (as specified below) to ensure consistent formatting and ease of use for analysis through other functions and sheets.
- Pre-test score and post-test score columns sum up total answers correct per test
- Survey_q1-4 records survey results, q1 is a numerical rating between 1-5 and q2-4 are yes/no questions encoded in binary

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Workshop**	**Date**		**Name**	**Age**		**pre-test score**	**post-test score**	**survey_q1**	**survey_q2**	**survey_q3**	**survey_q4**
2												
3	Text	=date(YYYY, mm,dd)		Text	Numerical		Numerical	Numerical	Numerical	Binary	Binary	Binary
4												

Workshop Data Sheet Contd.

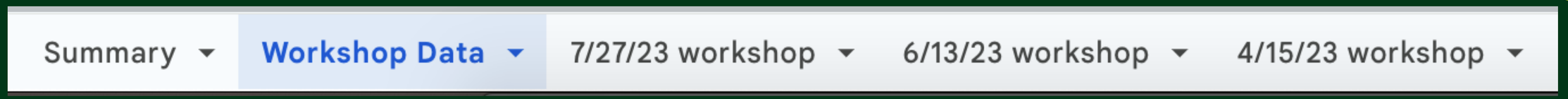
	A	B	C	D	E	F	G	H	I	J	K	L
1	**Workshop**	**Date**		**Name**	**Age**		**pre-test score**	**post-test score**	**survey_q1**	**survey_q2**	**survey_q3**	**survey_q4**
2	RP	7/27/23		Emily Johnson	18		5	5	5	1	1	1
3	RP	7/27/23		Daniel Martinez	16		#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
4	RP	7/27/23		Sophia Williams	17		3	5	5	1	1	1
5	RP	7/27/23		Liam Thompson	28		2	5	4	1	1	1
6	RP	7/27/23		Olivia Garcia	16		1	5	5	1	1	1
7	RP	7/27/23		Ethan Smith	17		0	4	5	0	1	1
8	RP	7/27/23		Ava Anderson	19		5	5	4	1	1	1
9	RP	7/27/23		Jackson Brown	20		4	#N/A	#N/A	#N/A	#N/A	#N/A
10	RP	7/27/23		Mia Taylor	#N/A		3	4	4	1	1	1
11	Other	4/15/23		Noah Davis	20		5	5	5	1	1	1
12	Other	4/15/23		Harper Wilson	21		4	5	5	1	0	1
13	Other	4/15/23		Aiden Rodriguez	22		3	5	5	1	1	1
14	Other	4/15/23		Isabella Martin	23		2	5	5	1	1	1
15	Other	4/15/23		Lucas Miller	18		#N/A	4	5	1	1	1
16	Other	4/15/23		Sophia Martinez	15		0	5	5	1	1	1
17	Other	4/15/23		Elijah Anderson	14		5	5	5	1	0	0

The above screenshot illustrates a hypothetical dataset, serving as a visual example of populated entries. It is important to note that empty responses are anticipated and handled through utilization of the =NA() function, resulting in the display of #N/A values as a means of addressing missing data points. This approach ensures a comprehensive and accurate handling of diverse scenarios within the dataset.

MM/DD/YY Workshop Sheet

This sheet is designed for duplication

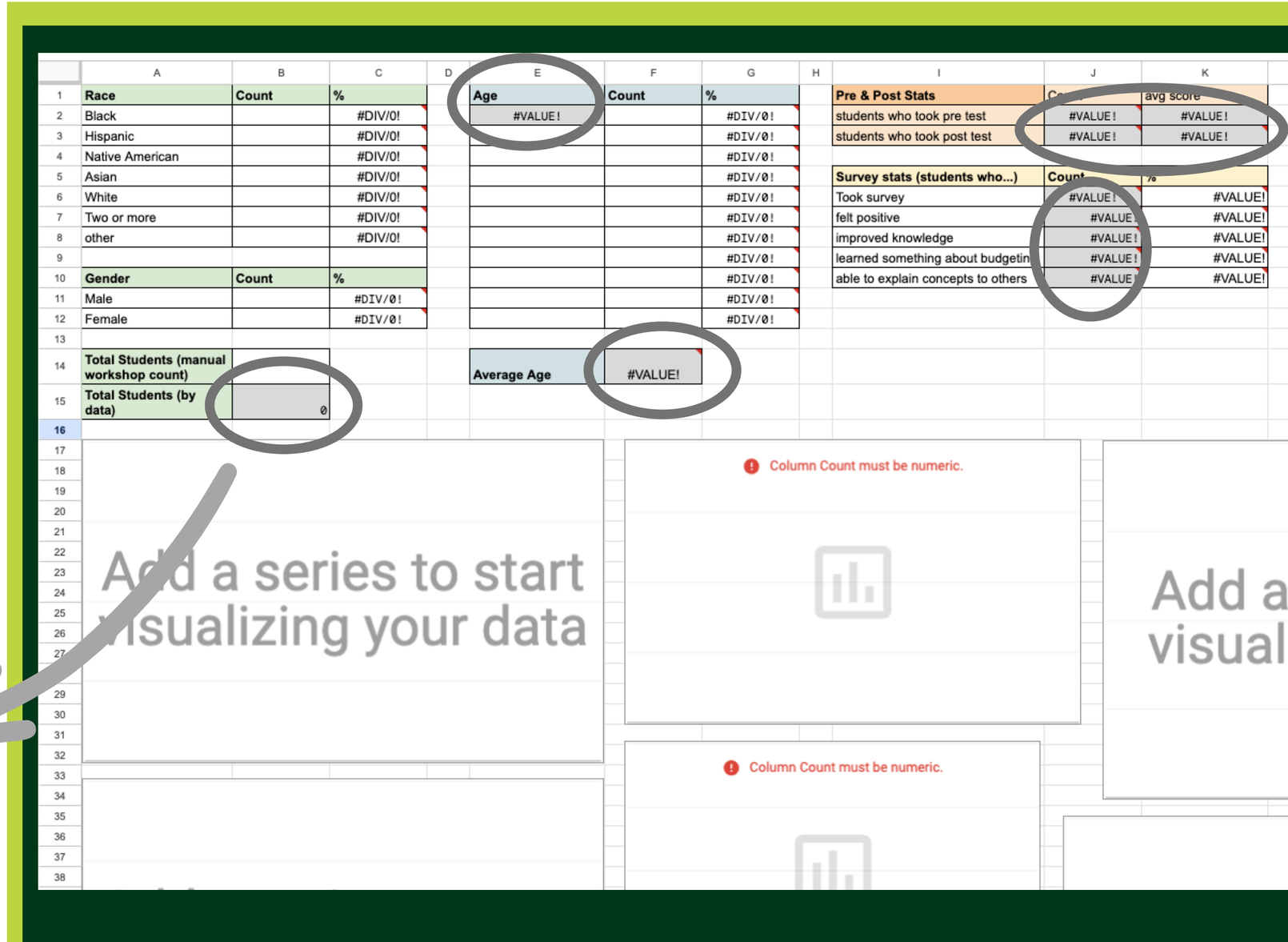
- Catering to individual workshops in order to extract unique insights from each session
- As all workshop data is consolidated within a single sheet (the "workshop data" sheet), differentiation between workshops is achieved through the use of the date column
- This sheet's title incorporates the date to signify this distinction
- The screenshot below demonstrates what tabs would look like for each individual workshop.



MM/DD/YY Workshop Sheet Contd.

Tables and charts are pre-populated in order to streamline and automate the process as much as possible. Upon initial review, the sheet will be blank but will start to update once the following two steps are completed:

1. Data entry into the "Workshop Data" sheet
2. Updating all functions in the (circled) cells that are grey with the corresponding workshop date.



MM/DD/YY Workshop Sheet Contd.

	A	B	C
1	**Race**	**Count**	**%**
2	Black		#DIV/0!
3	Hispanic		#DIV/0!
4	Native American		#DIV/0!
5	Asian		#DIV/0!
6	White		#DIV/0!
7	Two or more		#DIV/0!
8	other		#DIV/0!
9			
10	**Gender**	**Count**	**%**
11	Male		#DIV/0!
12	Female		#DIV/0!
13			
14	**Total Students (manual workshop count)**		
15	**Total Students (by data)**	0	

=COUNTIF('Workshop Data'!B:B, "=**mm/dd/yy**")

The count figures in the race and gender tables are inputted manually, derived from visual observations during workshops.

To ascertain the total number of students, an option is provided, allowing for either manual entry or data-based calculation, based on preference and accuracy. (All other calculations use the data-based total student count).

% columns are calculated by dividing the count by total students and rounding for simplicity.

(Hypothetical data)

	A	B	C
1	**Race**	**Count**	**%**
2	Black	7	30.43
3	Hispanic	7	30.43
4	Native American	2	8.7
5	Asian	4	17.39
6	White	1	4.35
7	Two or more	1	4.35
8	other	1	4.35
9			
10	**Gender**	**Count**	**%**
11	Male	13	56.52
12	Female	10	43.48
13			
14	**Total Students (manual workshop count)**	23	
15	**Total Students (by data)**	23	

=round(B12/B15%, 2)

=COUNTIF('Workshop Data'!B:B, "=**07/13/23**")

MM/DD/YY Workshop Sheet Contd.

The Age table is constructed using a Google API Query function:

```
=transpose(query('Workshop Data'!$A:$L, "select count(D) where B = date 'yyyy-mm-dd' pivot E ",-1))
```

The query is similar to SQL, and this specific function pivots the age column to extract the unique ages present in the workshop and calculates the count of each unique age. Null represents an attendee who wrote their name, but left the age section of the form blank.
Average age is calculated utilizing a query as well by filtering for the respective workshop date and extracting the average of all ages from the age column:

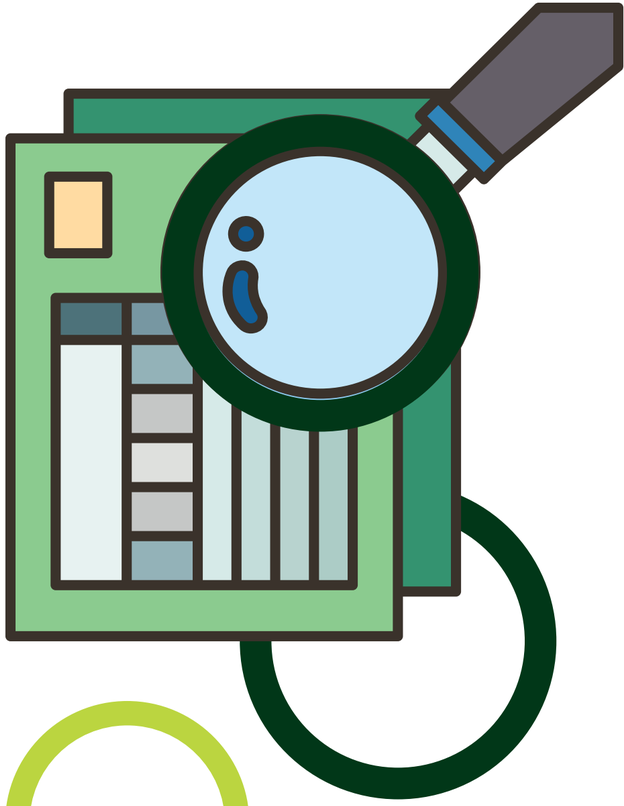
```
=round(transpose(query('Workshop Data'!$A:$L, "select avg(E) where B = date 'yyyy-mm-dd' label avg(E) "", -1)),1)
```

(The bold date section in each query is what must be updated for each workshop sheet.)

(Hypothetical data)

E	F	G
Age	**Count**	**%**
null	1	4.35
16	4	17.39
17	4	17.39
18	3	13.04
19	3	13.04
20	1	4.35
21	1	4.35
22	3	13.04
23	1	4.35
28	1	4.35
31	1	4.35
Average Age	19.6	

MM/DD/YY Workshop Sheet Contd.



The following tables also make use of multiple Google API Queries. Notably, the queries incorporate a 'where' condition to ensure that the pre-test and post-test columns are not null, thus accurately reflecting students who undertook both assessments. The query below provides the pre test student count, all other queries for this table are structurally analogous, involving minor adjustments such as selecting different columns (pretest score column vs. posttest score column) or employing 'avg' instead of 'count' for computations.

I	J	K
Pre & Post Stats (Only counts students who took both pre and post test)	Count	avg score
students who took pre test	20	2.55
students who took post test	20	4.25

=QUERY('Workshop Data'!$A:$L, "Select count(G) where B = date '**yyyy-mm-dd**' and (G is not null and H is not null) label count(G) ''")

MM/DD/YY Workshop Sheet Contd.

Finally, the survey stats table also makes use of queries, utilizing various logical statements.

=QUERY('Workshop Data'!$A:$L, "Select count(B) where B = date '**yyyy-mm-dd**' and (I is not null or J is not null or K is not null or L is not null) label count(B) '"')

The logical statement in this case not only filters by date but ensures that at least 1 of the four survey questions is answered.

=QUERY('Workshop Data'!$A:$L, "Select count(I) where B = date '**yyyy-mm-dd**' and (I >= 4) label count(I) '"')

A rating of 4 or 5 demonstrates positive feelings towards the workshops.

The use of binary for survey questions 2-4 allow for simple extraction of 'yes' answers through logical statements that collect count of ones.

Survey stats (students who...)	Count	%
Took survey	20	86.96
felt positive	18	90
improved knowledge	16	80
learned something about budgeting	17	85
able to explain concepts to others	20	100

=QUERY('Workshop Data'!$A:$L, "Select count(J) where B = date '**yyyy-mm-dd**' and (J = 1) label count(J) '"')

=QUERY('Workshop Data'!$A:$L, "Select count(K) where B = date '**yyyy-mm-dd**' and (K =1) label count(K) '"')

=QUERY('Workshop Data'!$A:$L, "Select count(L) where B = date '**yyyy-mm-dd**' and (L = 1) label count(L) '"')

MM/DD/YY Workshop Sheet Contd.

Upon completion of all table entries, the charts and visual representations will seamlessly auto-generate. No further action is needed to ensure that these charts are complete. A total of seven distinct visualizations have been pre-configured to correspond with specific tables. There are pie charts created for the age, race, and gender distributions. Race and gender are also visualized using a bar graph in order to provide better comparability between each characteristic. A bar graph is used to demonstrate pre to post test score improvement, and another which lists out survey result percentages.

The automation of both tables and charts allow for for seamless integration and consistency for each individual workshop analysis.

An overview of a completed mm/dd/yy workshop sheet (hypothetical data)

Race	Count	%
Black	7	30.43
Hispanic	7	30.43
Native American	2	8.7
Asian	4	17.39
White	1	4.35
Two or more	1	4.35
other	1	4.35

Gender	Count	%
Male	13	56.52
Female	10	43.48

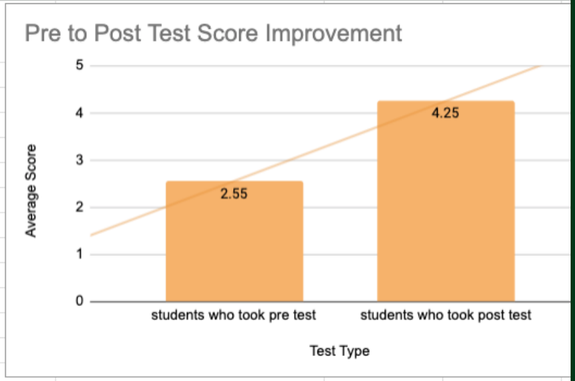
Total Students (manual workshop count)	23
Total Students (by data)	23

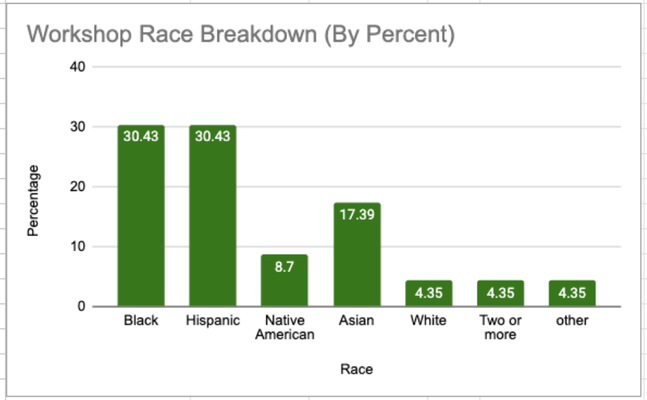
Age	Count	%
null	1	4.35
16	4	17.39
17	4	17.39
18	3	13.04
19	3	13.04
20	1	4.35
21	1	4.35
22	3	13.04
23	1	4.35
28	1	4.35
31	1	4.35

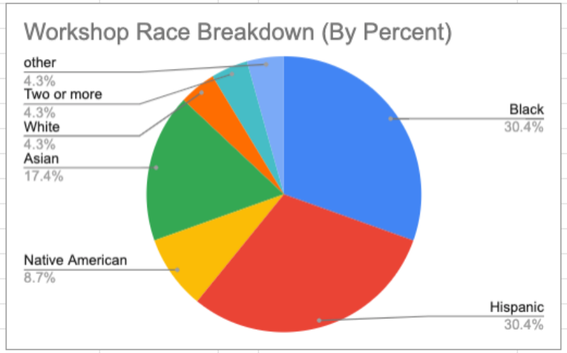
Average Age	19.6

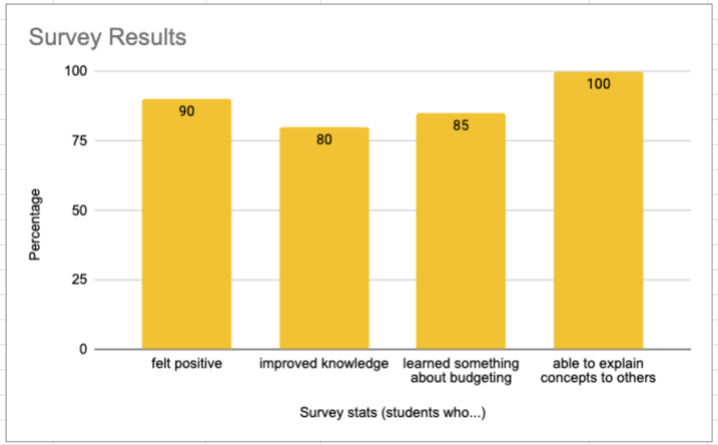
Pre & Post Stats (Only counts students who took both pre and post test)	Count	avg score
students who took pre test	20	2.55
students who took post test	20	4.25

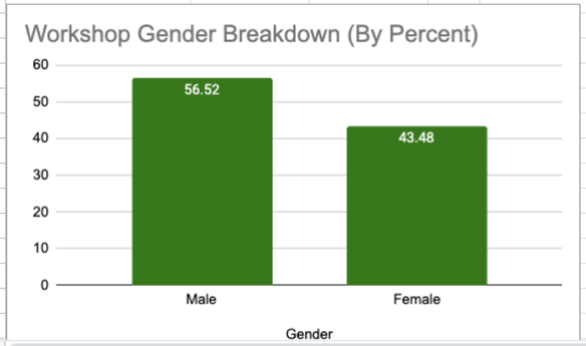
Survey stats (students who...)	**Count**	**%**
Took survey	20	86.96
felt positive	18	90
improved knowledge	16	80
learned something about budgeting	17	85
able to explain concepts to others	20	100

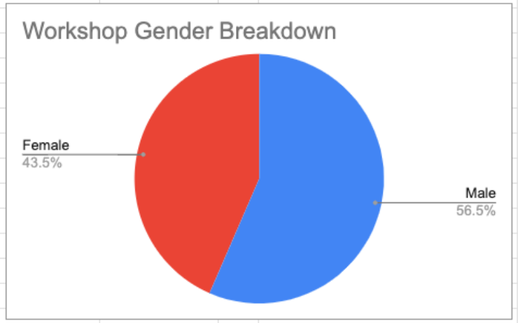


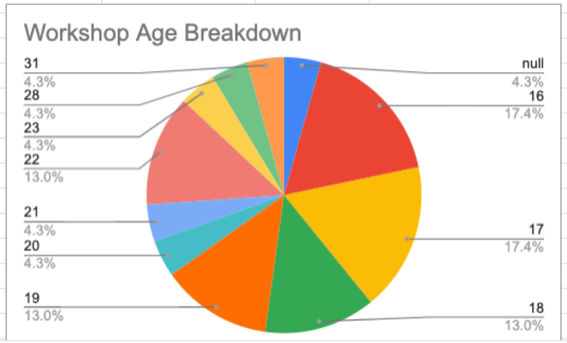






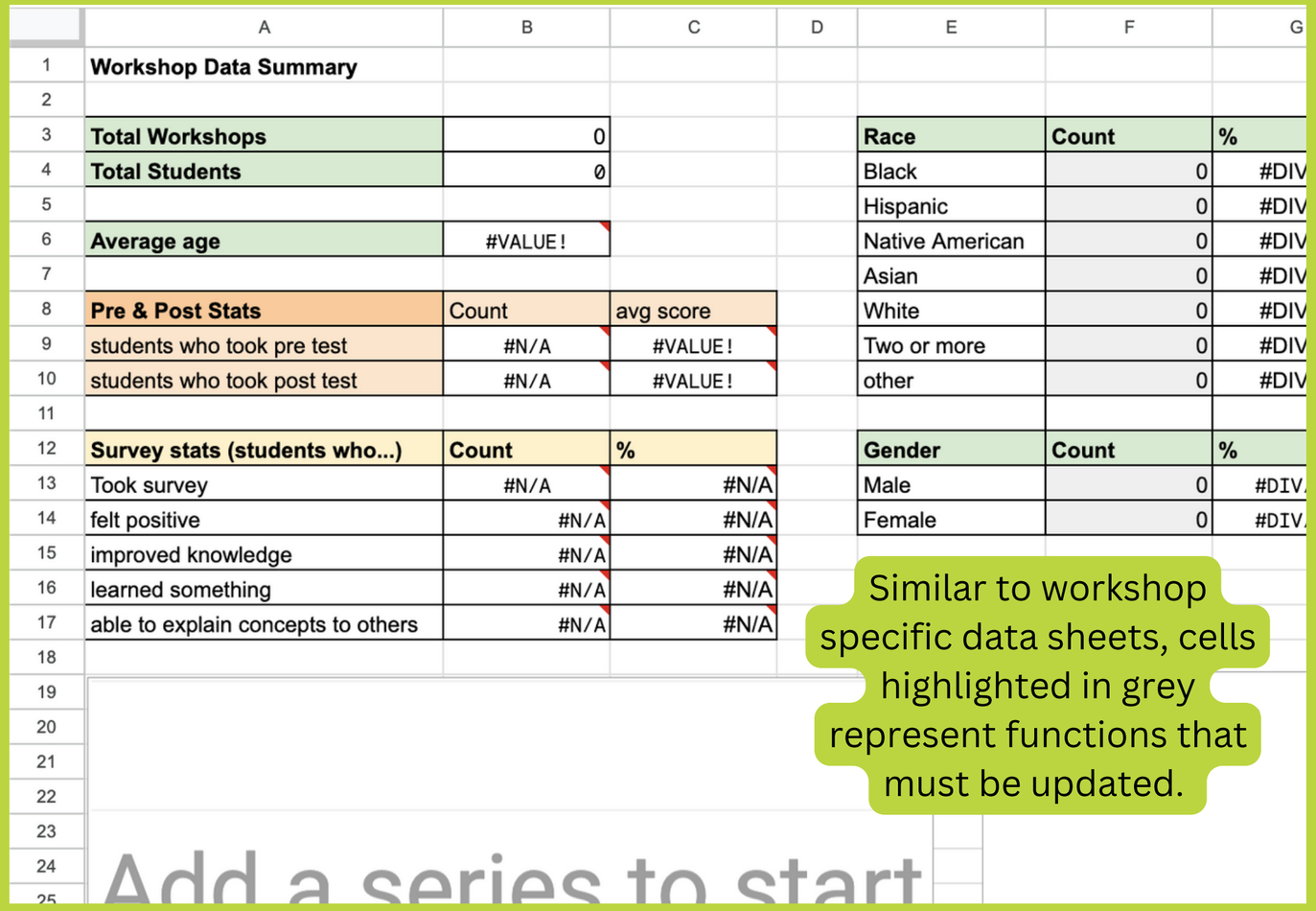






Summary Sheet

The summary sheet is a simplified version of the "MM/DD/YY Workshop" sheet but aggregates the entirety of all workshop data. This sheet is meant to be utilized to get a generalized view for all events held in a certain time period (annual, quarterly, etc). Functions have been inserted to calculate total workshops held so far, total students who have attended, and average age of students. There are also pre and post-test statistics, survey results, as well as race and gender breakdown tables similar to workshop specific sheets. This synthesis streamlines data analysis, fostering a comprehensive understanding of workshop trends and outcomes.



	A	B	C	D	E	F	G
1	**Workshop Data Summary**						
2							
3	**Total Workshops**	0			**Race**	**Count**	**%**
4	**Total Students**	0			Black	0	#DIV
5					Hispanic	0	#DIV
6	**Average age**	#VALUE!			Native American	0	#DIV
7					Asian	0	#DIV
8	**Pre & Post Stats**	Count	avg score		White	0	#DIV
9	students who took pre test	#N/A	#VALUE!		Two or more	0	#DIV
10	students who took post test	#N/A	#VALUE!		other	0	#DIV
11							
12	**Survey stats (students who...)**	**Count**	**%**		**Gender**	**Count**	**%**
13	Took survey	#N/A	#N/A		Male	0	#DIV
14	felt positive	#N/A	#N/A		Female	0	#DIV
15	improved knowledge	#N/A	#N/A				
16	learned something	#N/A	#N/A				
17	able to explain concepts to others	#N/A	#N/A				
18							
19							
20							
21							
22							
23							
24							
25							

Add a series to start

Similar to workshop specific data sheets, cells highlighted in grey represent functions that must be updated.

Summary Sheet Contd.

	A	B	C
1	**Workshop Data Summary**		
2			
3	**Total Workshops**	4	
4	**Total Students**	72	
5			
6	**Average age**	19	
7			
8	**Pre & Post Stats**	Count	avg score
9	students who took pre test	66	2.424242424
10	students who took post test	66	4.378787879
11			
12	**Survey stats (students who...)**	**Count**	**%**
13	Took survey	68	94.44
14	felt positive	58	80.56
15	improved knowledge	49	68.06
16	learned something	48	66.67
17	able to explain concepts to others	56	77.78

=COUNTUNIQUE('Workshop Data'!B2:B)

=COUNT('Workshop Data'!B2:B)

=round(query('Workshop Data'!$A:$L, "select avg(E) label avg(E) ''"),1)

=round(query('Workshop Data'!$A:$L, "select avg(E) label avg(E) ''"),1)

The queries for the pre and post stats and survey results are analogous to the workshop specific queries as explained earlier, excluding date as a condition.

(Hypothetical data)

Summary Sheet Contd.

Race	Count	%
Black	15	20.83
Hispanic	16	22.22
Native American	7	9.72
Asian	9	12.5
White	3	4.17
Two or more	3	4.17
other	2	2.78
Gender	**Count**	**%**
Male	29	40.28
Female	26	36.11

Race and gender data is not included in the "workshop data" sheet, and since it is difficult to pre-determine how many workshop specific tabs will be created along with their dates, these tables have two options:

1. Manually sum and input data
2. Utilize =SUM(x,y,z) function where each variable references the count per row, per workshop specific sheet
3. Utilize =sum('mm/dd/yy workshop'!B2,'mm/dd/yy workshop'!B2,....) as the most automated method and update date and number of workshops in the sum functions accordingly

An overview of a completed summary workshop sheet (hypothetical data)

Workshop Data Summary

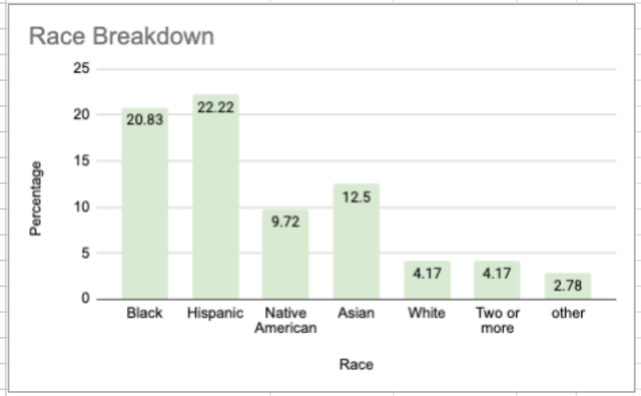
Total Workshops	4
Total Students	72
Average age	19

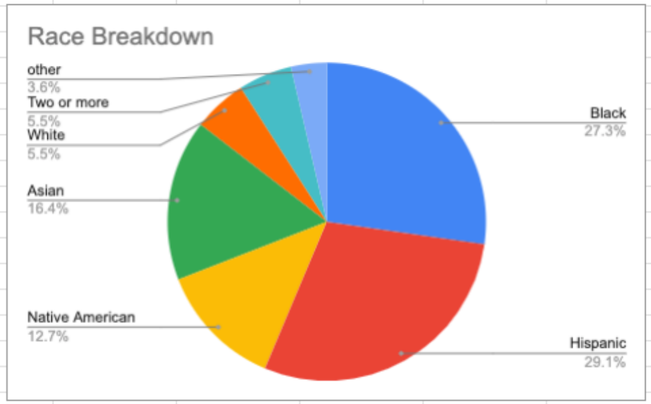
Pre & Post Stats	Count	avg score
students who took pre test	66	2.424242424
students who took post test	66	4.378787879

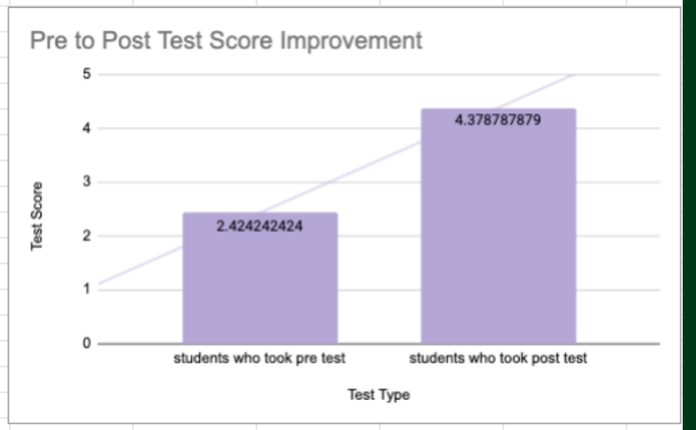
Survey stats (students who...)	**Count**	**%**
Took survey	68	94.44
felt positive	58	80.56
improved knowledge	49	68.06
learned something	48	66.67
able to explain concepts to others	56	77.78

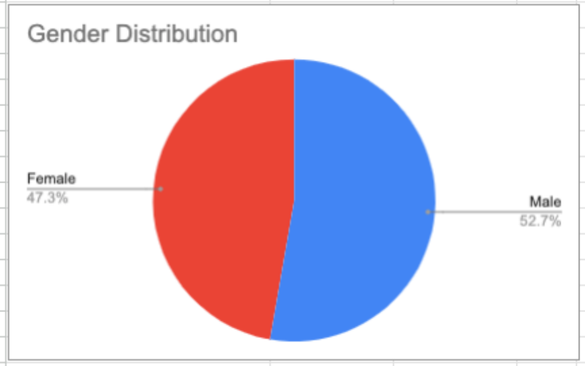
Race	**Count**	**%**
Black	15	20.83
Hispanic	16	22.22
Native American	7	9.72
Asian	9	12.5
White	3	4.17
Two or more	3	4.17
other	2	2.78

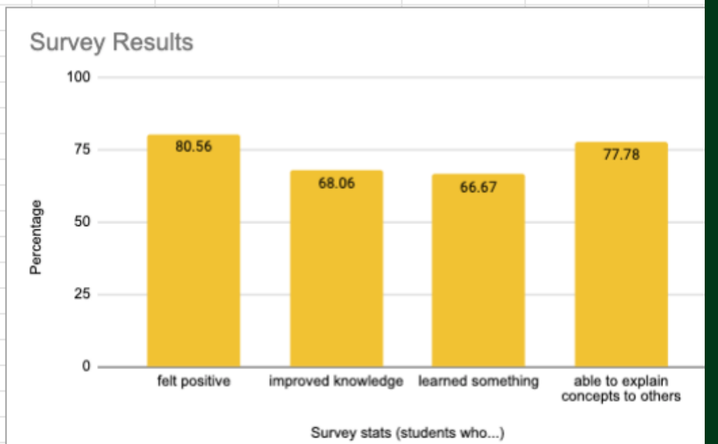
Gender	**Count**	**%**
Male	29	40.28
Female	26	36.11











Conclusion

Through this comprehensive, updated data review process, YFIA aims to consistently store workshop data, and extract insights on attendee demographics to continue to better serve the community. Separated sheets inside the main spreadsheet help maintain a sense of structure and aid in evaluating workshops by topic, organization, and other metrics. The plan currently set in place is to create new master spreadsheets annually in order to ensure that the data is manageable. Given the relatively modest scale of data for annual workshops, the need for an enterprise data platform is presently unnecessary but potential adoption in the future remains open for consideration. The implemented process eliminates the requirement for manual value calculations, given the extensive automation of functions and automatic visualization generation enhances review time.

Thank you!