Michael Moore, MS

2023 March 16

brief data analysis & visualization portfolio ~

1. Organizing and analyzing data

Quantitative ~

With organizing and analyzing quantitative data,

my first choice is to work in RStudio with R and RMarkdown.

This provides the opportunity to very carefully document the work

by putting it into the programmatic approach of R

while adding additional description in narrative around the coding.

It becomes very easy to review and check the work, as well as carefully make changes if needed.

And it naturally provides a detailed template for repeating similar work in the future.

For analysis, in this sample I only show some simple non-parametric tests.

This is an important and common scenario, where some simple and non-random data may not support other quantitative analysis.
It also fits the space here well ~

With my Applied Statistics graduate certificate work,

I am also comfortable with justifying and performing parametric hypothesis tests, advanced linear regression, and ANOVA in R,

as well as several exploratory multivariate techniques given the right software such as JMP.

Load data

Hide

```
### load the given data file


#--- use package openxlsx for easy loading from excel , including by named range ---
suppressPackageStartupMessages(library(openxlsx))

#--- pull by a named range precisely defining the data cells ---
tryCatch(
          {
              dataRootPath = "./data/"
              dataFullPath <- paste(dataRootPath, "test dataset.xlsx", sep="")

              dataInDF <- openxlsx::read.xlsx(dataFullPath, namedRegion="allData")
          },
          warning=function(cond) {
              stop( cond )
          }
       )
skimr::skim(dataInDF)
```

Table 1: Data summary

Name	dataInDF
Number of rows	100
Number of columns	4
Column type frequency:	
numeric	4
Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
ID	0	1.00	50.50	29.01	1	25.75	50.5	75.25	100	▇▇▇▇▇
ORG	3	0.97	1.82	0.69	1	1.00	2.0	2.00	3	▇▁▇▁▂
PRE_P	6	0.94	7.99	5.42	0	4.00	7.0	12.00	22	▇▅▃▃▂
POST_P	3	0.97	12.41	4.05	0	9.00	12.0	15.00	21	▁▂▅▇▃

clean missing data, initially going with dropping

While it might be interesting to make a case for some kind of imputation,
I'm not sure that there really is one here.

With the very simple calculations anticipated, there just isn't much need for it in the first place.
We'll mostly look at, I think, the median pre-post delta for each group. Imputing the median should not change anything
More fancy techniques don't makes sense with so few variables.

Hide

```
org1onlyDF <- dplyr::filter(dataInDF, ORG==1)
org2onlyDF <- dplyr::filter(dataInDF, ORG==2)
org3onlyDF <- dplyr::filter(dataInDF, ORG==3)

org1_missingData <- org1onlyDF[!complete.cases(org1onlyDF), ]
org1_completeCases <- org1onlyDF[complete.cases(org1onlyDF), ]

org2_missingData <- org2onlyDF[!complete.cases(org2onlyDF), ]
org2_completeCases <- org2onlyDF[complete.cases(org2onlyDF), ]

org3_missingData <- org3onlyDF[!complete.cases(org3onlyDF), ]
org3_completeCases <- org3onlyDF[complete.cases(org3onlyDF), ]


noOrgDF <- dplyr::filter(dataInDF, !(ORG %in% c(1,2,3)))

#---
allCompleteCasesDF <- rbind(org1_completeCases, org2_completeCases, org3_completeCases)
```

Dropping incomplete cases for all gives us:

Org	# complete cases	# dropped rows	(of total)
1	29	4	33
2	43	5	48
3	16	0	16

and with 3 rows dropped for having no Org.

Also clean a couple out-of-bounds values

I think we must also drop the values with out-of-bounds Pre- or Post-Test values.
There is a 22 and a 21 in there.
That is clearly incorrect with a max of 20, and I don't know a way to logically correct it.

Hide

```
allCompleteAndInBoundsCasesDF <- dplyr::filter(allCompleteCasesDF, !(PRE_P > 20))
allCompleteAndInBoundsCasesDF <- dplyr::filter(allCompleteAndInBoundsCasesDF, !(PRE_P < 0))

allCompleteAndInBoundsCasesDF <- dplyr::filter(allCompleteAndInBoundsCasesDF, !(POST_P > 20))
allCompleteAndInBoundsCasesDF <- dplyr::filter(allCompleteAndInBoundsCasesDF, !(POST_P < 0))
```

So what do we have for final analysis and visualization?

Let's pull all the interesting values explicitly to be super clear:

Hide

```
numberOrg1fullyCleaned <- NROW(allCompleteAndInBoundsCasesDF[allCompleteAndInBoundsCasesDF$ORG==1,])
numberDroppedOrg1forMissing <- NROW(org1_missingData)
numberDroppedOrg1forOutOfBounds <- (NROW(org1_completeCases) - numberOrg1fullyCleaned)
numberOrg1raw <- NROW(org1onlyDF)

numberOrg2fullyCleaned <- NROW(allCompleteAndInBoundsCasesDF[allCompleteAndInBoundsCasesDF$ORG==2,])
numberDroppedOrg2forMissing <- NROW(org2_missingData)
numberDroppedOrg2forOutOfBounds <- (NROW(org2_completeCases) - numberOrg2fullyCleaned)
numberOrg2raw <- NROW(org2onlyDF)

numberOrg3fullyCleaned <- NROW(allCompleteAndInBoundsCasesDF[allCompleteAndInBoundsCasesDF$ORG==3,])
numberDroppedOrg3forMissing <- NROW(org3_missingData)
numberDroppedOrg3forOutOfBounds <- (NROW(org3_completeCases) - numberOrg3fullyCleaned)
numberOrg3raw <- NROW(org3onlyDF)

numberAllOrgsFullyCleaned <- NROW(allCompleteAndInBoundsCasesDF)
numberAllOrgsDroppedForMissing <- (numberDroppedOrg1forMissing + numberDroppedOrg2forMissing + numberDroppedOrg3
numberAllOrgsDroppedForOutOfBounds <- (numberDroppedOrg1forOutOfBounds + numberDroppedOrg2forOutOfBounds + number
numberAllOrgsRaw <- (numberOrg1raw + numberOrg2raw + numberOrg3raw)
```

Org	# fully cleaned cases	# dropped missing	# dropped out-of-bounds	(of # raw)
1	28	4	1	33
2	43	5	0	48
3	15	0	1	16
total	**86**	**9**	**2**	**97**

remembering that 3 rows had no Org # in the first place

Get some stats on median Pre- and Post-Test scores

This simple data is a list of Pre- and Post-Test scores alone.

It is important to match the statistics to the type of data we have and they type of analysis it can support.
With this very simple data of relatively unknown origin, let's start with a non-parametric test -

Hide

```
#--- let's also do a minimal check that the distributions have "statistically significant" difference
#       I choose a non-parametric test for multiple medians

kwResults <- stats::kruskal.test(POST_P ~ ORG, data=allCompleteAndInBoundsCasesDF)
kwResultsPval_out = paste("Kruskal-Wallis p-value is: ", kwResults$p.value, sep="")
cat(kwResultsPval_out)

rstatix::dunn_test(allCompleteAndInBoundsCasesDF, POST_P ~ ORG, p.adjust.method = "bonferroni")
```

```
Kruskal-Wallis p-value is: 0.000000029002085838834
```

.y. <chr>	group1 <chr>	group2 <chr>	n1 <int>	n2 <int>	statistic <dbl>	p <dbl>
1 POST_P	1	2	28	43	-0.5308494	0.595523169434
2 POST_P	1	3	28	15	4.9671481	0.000000679447
3 POST_P	2	3	43	15	5.7299674	0.000000010045

3 rows | 1-8 of 10 columns

Qualitative ~

With organizing and analyzing qualitative data,

I am comfortable using a variety of tools for categorical and connecting analysis.

My first choice for organizing materials of all types – for qualitative analysis or other purposes – is Zotero.

For categorical analysis,

I have particularly used Dedoose and QDA Miner Lite, as well as Excel and also OneNote.

I most often have used a straightforward, phronetic approach such as advocated by Tracy, and am comfortable with flexible, inductive coding as well as deductive coding with a team.

I excel at both reductive coding and identifying cross-cutting themes.

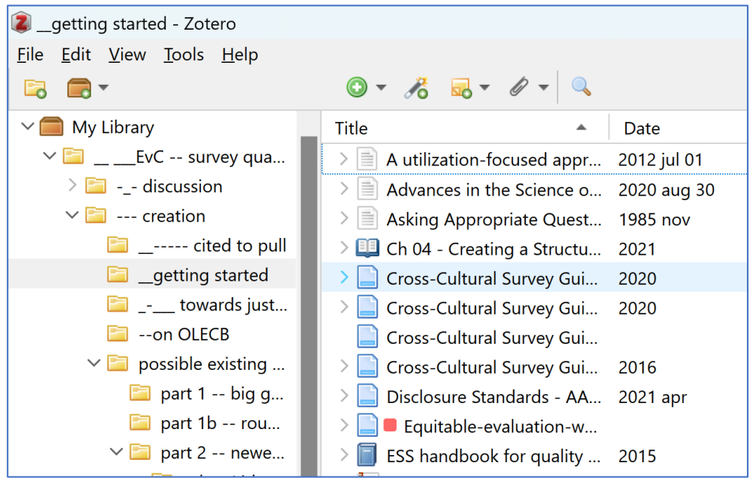
For connecting analysis,

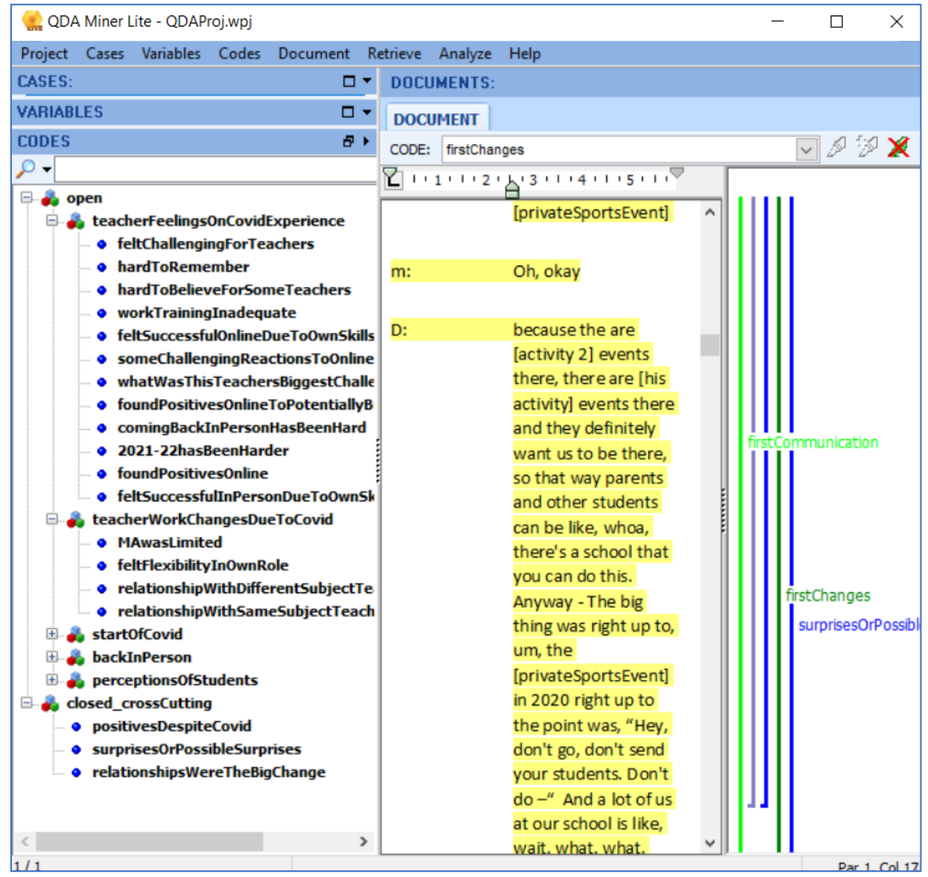
I have especially used OneNote,

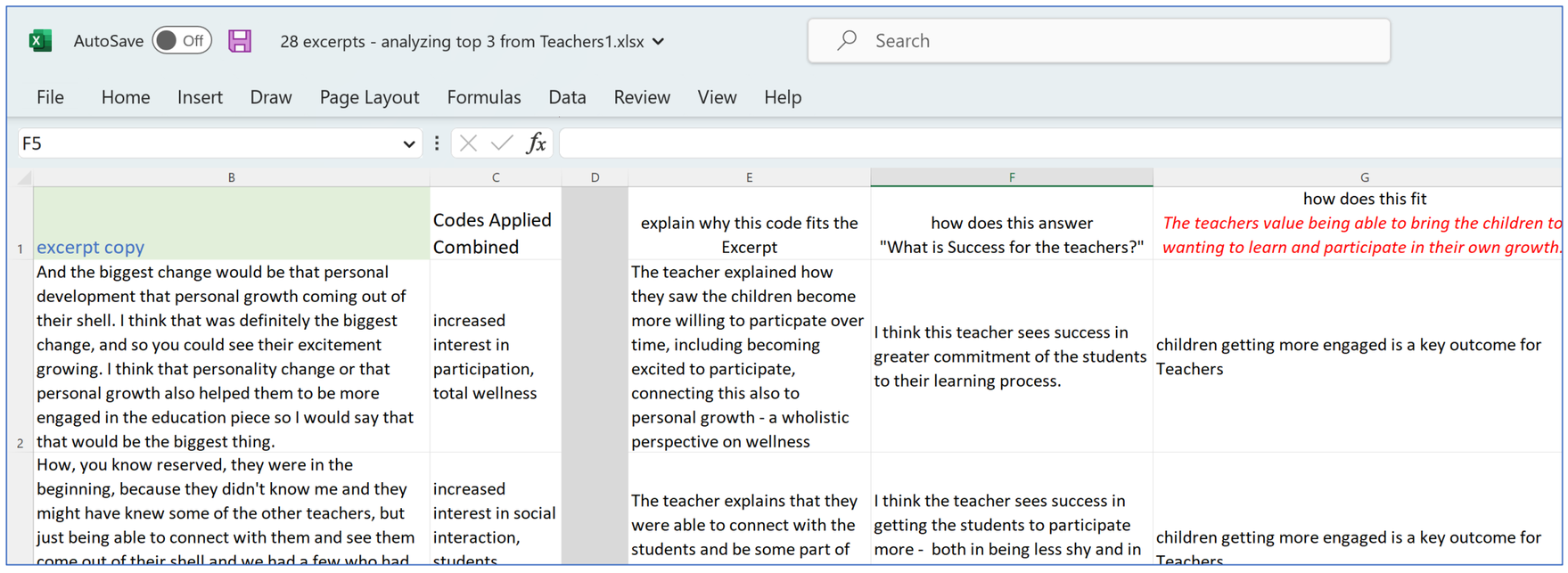
as well as, of course, exploring ideas through writing in any tool, such as OneNote or Word or any typing program.

My natural talent for synthesis makes me very effective in this type of work.

Because of its popularity and to demonstrate the more specific tools involved, I include images of categorical work.







2. Summarizing and visualizing data

Quantitative ~

With summarizing quantitative data,

I enjoy particularly working in R in RStudio where I can work most directly from data preparation there.

It also provides the opportunity to take advantage of a seemingly unlimited array of possibilities – most any data presentation one has seen can be created in R, given the effort.

I am also very comfortable with Excel, which can be helpful to decision-makers who appreciate the chance to engage with the source data in Excel tables driving the visualization.

As sample images,

I show at the top a highly customized dotplot in R that is enhanced with Word and PowerPoint.

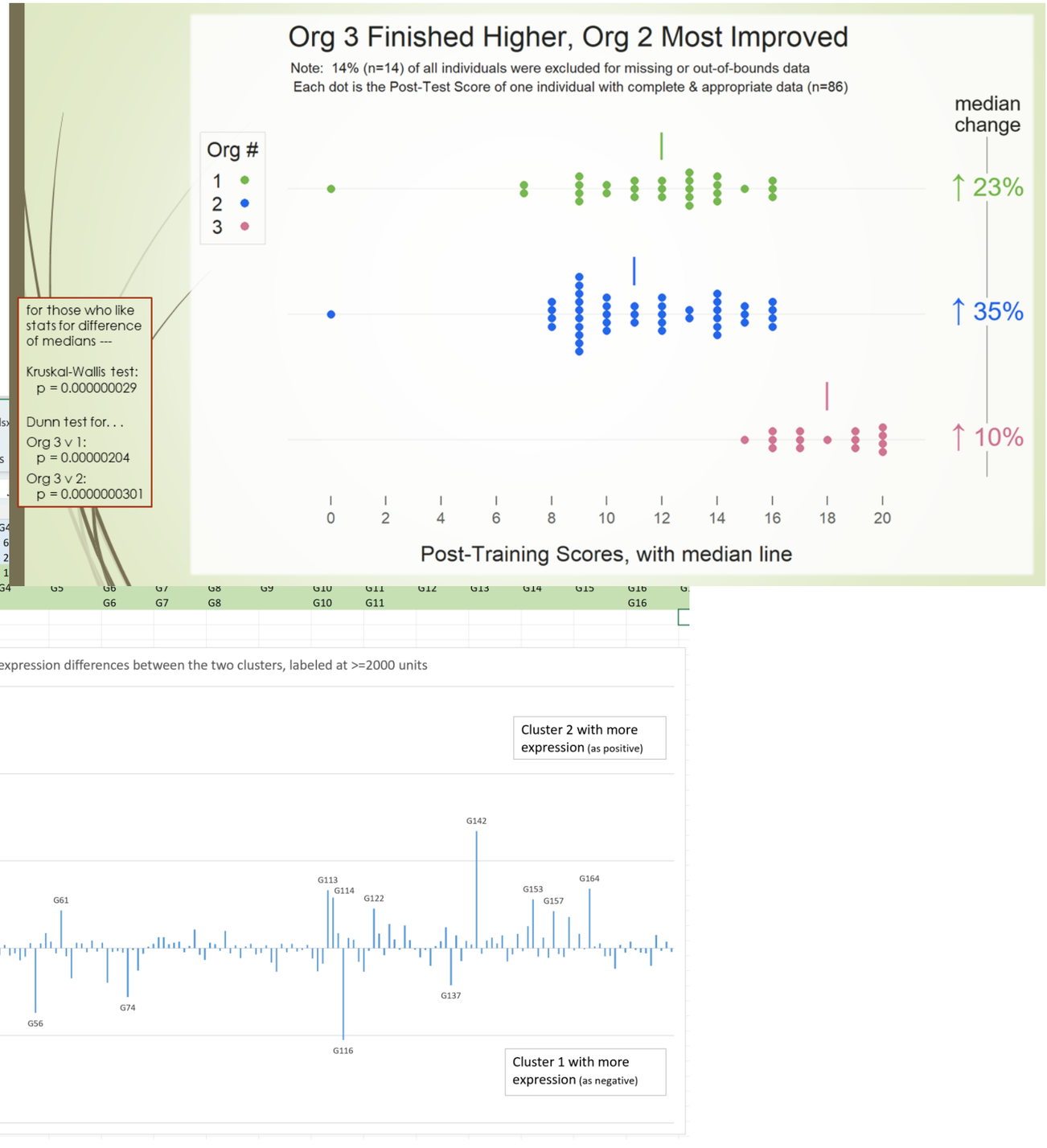
I chose a dotplot very deliberately to communicate results and to provoke meaningful next questions.

I added a secondary visualization of "median change" at the right as graphic elements in Word.

Then in PowerPoint, for a presentation, I also included some "hard stats" at the left.

Below is a chart in Excel used to illustrate the importance of many variables to two outcomes, split as up and down.

I label bars conditionally to highlight variables that reach a set level of interest.



Qualitative ~

With summarizing qualitative data,

I am comfortable exploring different types of visualizations, which are more limited with qualitative but do exist.

My comfort with R in RStudio and with Excel gives me access to most anything I am exposed to.

In my work to date, I have most often discussed categorical themes and connecting themes in narrative,
highlighting the conclusion, the explanation, and supporting text
in different ways.

As sample images,

I include a typical "word cloud" to show my comfort with this, made from words in a version of my resumé

as well as two examples of narrative summaries from real-world projects.

The first narrative summary provides conclusions next to big number icons, with evidence described in complete sentences and references to endnotes for the source behind it.

The second gives conclusions in bullet points for one of several groups, highlighted in large text – "Parents" – with evidence included directly below in top quotes from that group.

Many outcomes may be stronger with art interventions

1 Political art has already been growing in importance in places that ICTJ works

Political art has already been growing in importance in many of the contexts where ICTJ works. Activists and protestors demanding for justice have increasingly used street art, photography, and other media to channel their demands, with amplified effect thanks to social media [69]. Examples of this locally-driven art activism include Lebanon, where youth-driven action using public art has built fresh engagement and connected with diverse groups across sectarian divides [70]. This interest suggests that using art interventions may be more successful by tapping into existing interest in such art.

2 By avoiding formal, technical language and processes, art interventions can be far more accessible, reaching much farther into civil society

By communicating in ways that people can appreciate without special training, art interventions can connect with regular people in an accessible way [71]. This includes connecting with populations that are difficult to reach, both in distant areas and in marginalized groups and the very old and young [72]. Using art also can open opportunities to work in traditional communication formats [73], create an even more targeted way to connect to people in ways they prefer and respect. Merritt-Cuneo[74] shared an important example of this last opportunity that came out of work in The Gambia, where certain womens groups that already hold a traditional role as communicators using songs and stories[75] became engaged in spreading understanding of transitional justice. Thes
areas that might otherwise not be exposed to outreach messages, and could co
and respected in those areas much more than formal, top-down messages likely

3 The emotional power of art interventions may make them
participants and society

The factors described in the previous section can also make for a safe
human rights abuse that transitional justice hopes to address. There are many
around their traumatic experiences did perceive such a safe space and this facili

Top elements of success for Parents

The parents we talked to were most impressed by...

- the variety of experiences in the program
- the high level of engagement that the activities created
- and the personal growth that their child or grandchild experienced

> *It really exceeded my expectation. It went over, because I was not expecting, you know, all the education and, like, all the hands-on activities for them to do to keep them busy throughout the day. I'm, I'm glad that we got to get in this program to really show us that it's - it's more out there.*
>
> *. . .my son. . .he's like a quiet kid and really didn't get out as much. As I put him in this program, you know, he started to talk a little more, and, you know - [he's] excited to go places like the drone [class]. . . he really liked that. He was like "yeah, I'm going!"*
>
> *...it's so much different than just learning from a textbook or watching a video. It was just in-depth, hands-on learning that expounded the curiosity in either the subject of space or even coding - and just to see their enthusiasm when they completed a game or completed a coding project -*
>
> *...it's kind of awakened an entrepreneurial spirit with her, so - and it's actually built her confidence to actually know what to do...I like the person that she's becoming and the young lady that she's developing into by stepping out of her comfort zone*

Read even more quotes for these top topics in Appendix A.