

April, 2017 - July 2017 Stats And Methods

Martin Gleason, MS

August 10, 2017

Purpose

This document is to summarize the last three months of CFO data, including: attendance policy, attendance numbers, referrals, JEMS issues, tracking and storing data, and data transformation. A way forward with CFOs will also be presented. While some of the tables maybe be difficult to read, thanks to the magic of PDF formats, any reader should be able to zoom into items of note. If the department wants to explore using R for stat and report generation, future editions of reports will be able to have interactive graphs.

Attendance

Critical to the continued success of the CFO is getting the attendance rate over 50%. Traditional JAC Orientation programming reached a cap of approximately 50%, with attendance numbers typically hovering between 25%-40%. CFOs run more frequently, in a variety of times and locations, and have additional attendance tracking problems. Solutions to these issue are presented in this report.

Attendance Policy

A carry over from traditional JAC programming, young people are referred to a CFO three times before not being eligible for the program. Where as traditional orientations did not have a referral process, CFOs do have a series of protocols inplace, including:

- Paper referrals that are given to each unit involved in CFO Programming
- Monthly lists (generated from JEMS based Excel Spreadsheets) that are given to SPOs of each unit
- Personalized form letters mailed between 5-10 days before a CFO
 - A copy of this letter is also given to POs
 - The letter's verbage reflects the more inclusive tone of the CFO, including "encouraging young people to attend" and their PO's last name as a person to contact if they cannot make the CFO
- Phone calls 12-48 hours before a CFO
 - All phone calls are documented in JEMS
- Attendance is recorded on paper and in JEMS
 - JEMS notes are typically entered within one week of a CFO
 - Excel is not consistently used to indicate attendance
 - This is more challenging than expected, but it will be explored further in the JEMS section

Attendance Numbers

Getting these numbers is problematic for a variety of reasons, not the least of which because JEMS does not track or export attendance numbers. To get the raw data for this particular report, the following was done:

1. JEMS exported 4 Excel tables
2. A new column, "Attendance" was added.
3. Paper records were consulted for determining attendance
4. Duplicate JEMS ID were removed through an Excel function

What the data indicates

The first ten youth from the new table are as follows

Table 1: First 10 CFO Youth in Alphabetical Order

FNAME	LNAME	Address	APRTNO	CITY	STATE	ZIP	SEX	JEMSID	POFNAME	POLNAME	CFO	Attended
Demarcus	Abram	13240 S Corliss	NA	Chicago	IL	60827	M	10326791	Joseph	Naegle	Y	N
Eric	Aguilar	10230 S Avenue G	NA	Chicago	IL	60617	M	10321465	Lisa	Bloechel	Y	Y
Juan	Alcaraz	1859 W Cullerton St	NA	Chicago	IL	60608	M	10316841	Nichollette	Vargas	N	Not Invited
Arquis	Alexander	4827 S Prairie Ave	NA	Chicago	IL	60621	M	10352192	Rodney	Purdy	Y	N
Xavier	Allen	10727 S Calumet Ave	1St Floor	Chicago	IL	60628	M	10342391	Victor	Junious	Y	N
Xavier	Allen	10727 S Calumet Ave	Unit 1	Chicago	IL	60628	M	10342391	Victor	Junious	Y	N
Latrell	Allen	5616 S Aberdeen St	NA	Chicago	IL	60621	M	10335541	Hannah	Siegel	Y	Y
Capria	Allen	5622 S Honore St	House	Chicago	IL	60636	F	10316032	Darryl	Dawkins	Y	Y
Julian	Almanza	2511 N Lawndale Ave	NA	Chicago	IL	60647	M	10341183	Antwan	Jones	N	Not Invited
Jose	Alvarez	3567 W Palmer Street	D	Chicago	IL	60647	M	10347745	Yvonne	Pulido	N	Not Invited

This first table illustrates some of the difficulties with the JEMS Data. Despite being “precleaned” in Excel, some errors remain. For example, this record:

Table 2: First Duplicate Youth

FNAME	LNAME	Address	APRTNO	CITY	STATE	ZIP	SEX	JEMSID	POFNAME	POLNAME	CFO	Attended
Xavier	Allen	10727 S Calumet Ave	1St Floor	Chicago	IL	60628	M	10342391	Victor	Junious	Y	N
Xavier	Allen	10727 S Calumet Ave	Unit 1	Chicago	IL	60628	M	10342391	Victor	Junious	Y	N

is duplicated, which inflates the number of youth counted by Excel.

Therefore, while it appears that the table has 134 young people who were invited to CFOs, the actual number is 122, a difference of 9.84%.

A 9.84% difference will probably not make or break a program. There is an even more difficult issue. Given that the bulk of the work on CFO attendance is done in Excel, these duplicate entries are also inflating attendance records. This kind of attendance error will complicate analysis and frustrate POs to no end. Fixing these duplication errors will be detailed in another section of this report.

Attendance Analysis

Earlier this month, the PYD unit was asked to compile a series of stats indicating attendance rates in CFOs. These numbers were painstakingly generated by handcounting rows of printed Excel tables. By loading this data into a different program a more thorough analysis can take place.

Cleaning

Before any sort of analysis can take place, duplicate CFO entries were eliminated. Only unique JEMSIDs were used in determining attendance data. Additionally, attendance codes were changed from a mix of cases to a consistent use of title case (each word except any articles were capitalized.) While this notation has its own set of issues; specifically, for entry into excel or JEMS, numbers would be faster and more consistent for attendance codes, this function collapsed codes like “in custody” into one term to be used across all CFO data. Additionally, all of the records went through this same function, to increase the aesthetic value of the report. Pretty reports are just more fun to read.

Tables and Graphs

Once the data was cleaned, additional attendance analysis was conducted. For example, the attendance records were broken into three distinct representations an aggregate total of attendance by code, attendance disaggregated by zip code, and two bar charts representing attendance numbers:

Table 3: Attendance of Youth (Aggregate)

Attended	Attendance_Count
N	89
Y	31
Saftey Issues	6
In Custody	5
In Adult Custody	1
Not Invited	1
Out Of State	1

This is a fast summary of CFO attendance numbers. The next chart separates the numbers into attendance by zipcode and gender.

Table 4: Attendance of Youth, by Zipcode

ZIP	Attended	Attendance Count by Zipcode
60617	N	5
60617	Y	7
60619	In Custody	1
60619	N	9
60619	Not Invited	1
60619	Y	7
60621	In Adult Custody	1
60621	In Custody	3
60621	N	13
60621	Out Of State	1
60621	Saftey Issues	1
60621	Y	4
60628	N	13
60628	Y	4
60636	In Custody	1
60636	N	9
60636	Saftey Issues	4
60636	Y	3
60644	N	11
60644	Y	2
60649	N	11
60649	Saftey Issues	1
60649	Y	3
60651	N	11
60651	Y	1
60827	N	7

Finally, two bar charts to show the same information graphically.

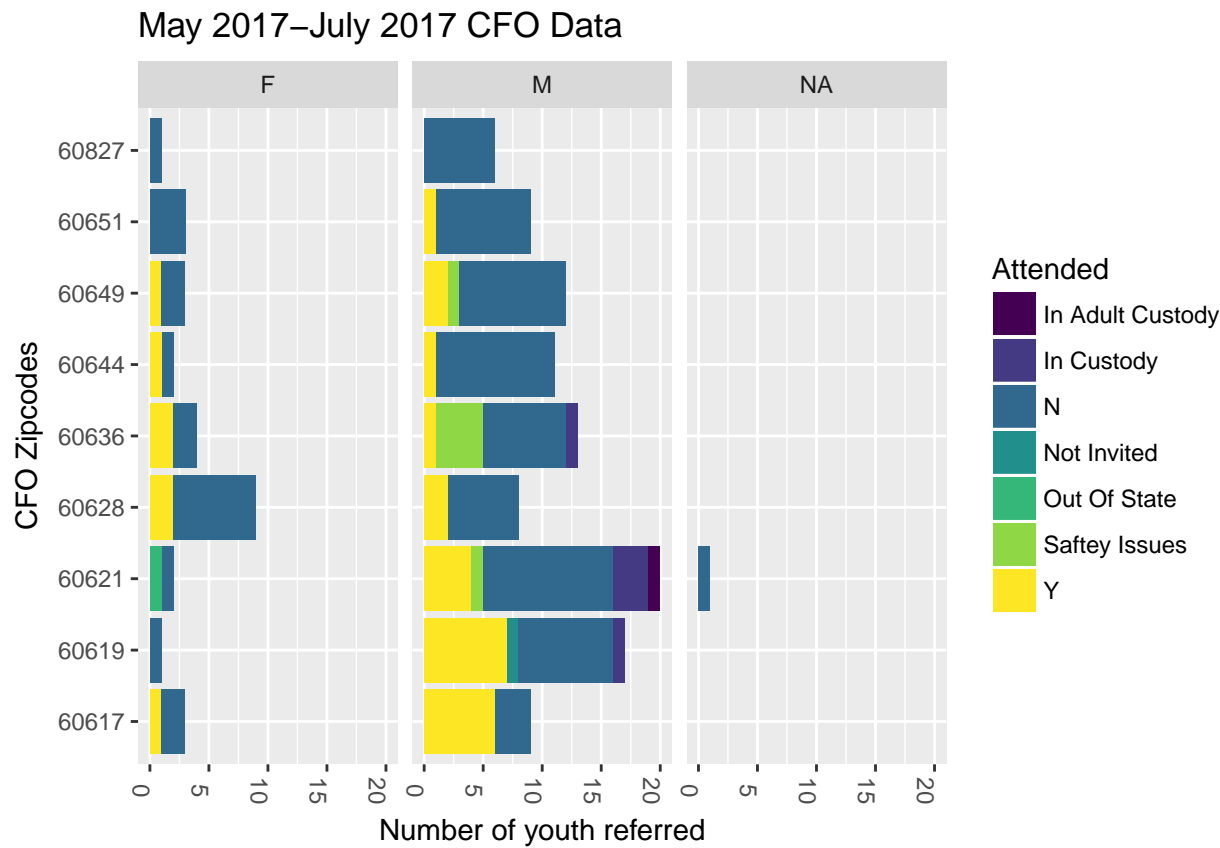


Figure 1: Attendance by Zip, Seperated by Gender

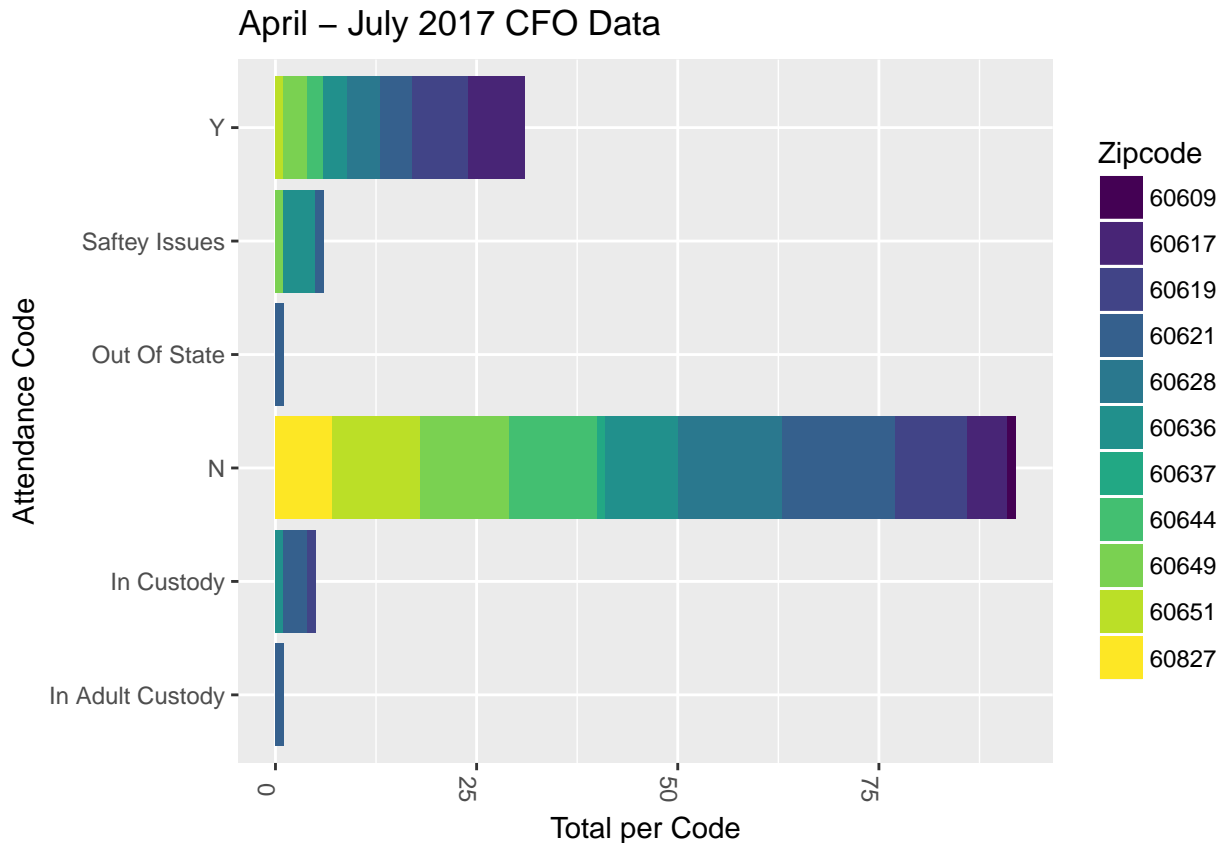


Figure 2: Attendance by Zip

With a few lines of code, two tables and two graphs were drawn to show how many young people have attended CFOs since April, 2017. This analysis also indicates the following issues:

1. The attendance categories of *N*, *Y*, *Not Invited*, *Safety Issues*, *Out Of State*, *In Custody*, *In Adult Custody* need clarification. Switching to numbers will improve clarity and analysis.
2. Some referrals do not have an assigned gender. This is due to human error, as these referrals appear to be added manually to the attendance charts.
 - JEMS needs to be updated to track non-binary youth.
3. The date of a CFO needs to be included on attendance forms or by JEMS queries.
4. Classifying CFO by Zipcode is a necessary evil. JEMS does not assign youth to CFO by police district, this is done manually. It is possible to have Excel group and assign by Zipcode, it is labor and intensive and not necessarily “push button repeatable.”

Even this raw table points out some key trends in CFO attendance. Of the 122 youth invited to attend a CFO since April, only 31 attended, a ratio of 25.4%. Given that this is an aggregate of all the CFOs since April, it would be helpful to break these numbers down further. Without dates, it is impossible to give attendance rates per CFO; however, these attendance stats can be broken down further by zipcode (see table 5.)

Table 5: Attendance per Zip, by Gender

ZIP	SEX	Attended	Attendance Rates
60617	F	N	2
60617	F	Y	1
60617	M	N	3
60617	M	Y	6
60619	F	N	1
60619	M	In Custody	1
60619	M	N	8
60619	M	Not Invited	1
60619	M	Y	7
60621	F	N	1
60621	F	Out Of State	1
60621	M	In Adult Custody	1
60621	M	In Custody	3
60621	M	N	11
60621	M	Saftey Issues	1
60621	M	Y	4
60621	NA	N	1
60628	F	N	7
60628	F	Y	2
60628	M	N	6
60628	M	Y	2
60636	F	N	2
60636	F	Y	2
60636	M	In Custody	1
60636	M	N	7
60636	M	Saftey Issues	4
60636	M	Y	1
60644	F	N	1
60644	F	Y	1
60644	M	N	10
60644	M	Y	1
60649	F	N	2
60649	F	Y	1
60649	M	N	9
60649	M	Saftey Issues	1
60649	M	Y	2
60651	F	N	3
60651	M	N	8
60651	M	Y	1
60827	F	N	1
60827	M	N	6

This data is the same as graph located at figure 1. It is seperated into a tabular format. This may make it more understandable for some; however, the sheer volume of information is overwhelming. It is also difficult to display in this format without additional softawre packages (future of editions of this report, if warranted, will address long tables.) Since it is possible to seperate the tables into more refined “chunks” of

information for ease of use, such tables should be created: All that is required are the parameters to observe this information.

Presenting this is limited only by the amount of data at hand. More data, such as CFO Dates, text mining of JEMS notes for key words, or any other data from JEMS would provide a wealth of information. An example of this, Given this particular data set, is attendance rates presented by gender, zipcode, and referring probation officer (see table six at the end of this document.)

Analysis

The tables and graphs suggest a relationship between PO, Zipcode, and Attendance. The nature of this relationship needs to be more thoroughly examined, especially by including dates, to determine if any of these factors can be tweaked to increase attendance. Given that the program is only three months old, these attendance hiccups can be addressed in a variety of methods. Evidence of the relationship (that we already assumed) provides PYD additional direction for improving attendance. As helpful as this is, more information about CFOs would be helpful in determining attendance issues; specifically, is the violence in Chicago's neighborhoods affecting CFO attendance?

Choropleth for Context

Drilling into the numbers by tables and graphs certainly helps paint the picture of who is, and is not, attending the CFOs, to answer this question, a map would be better. Given that it is possible to map the young people's addresses on file, with their attendance records, this may shed additional light on the attendance figures. Adding interactive (zoom-able) maps into a PDF is possible for future editions of this report, as is refining crime data to show case specific kinds of crime. But as of this writing, it is possible to plot referral addresses on a map of Chicago.

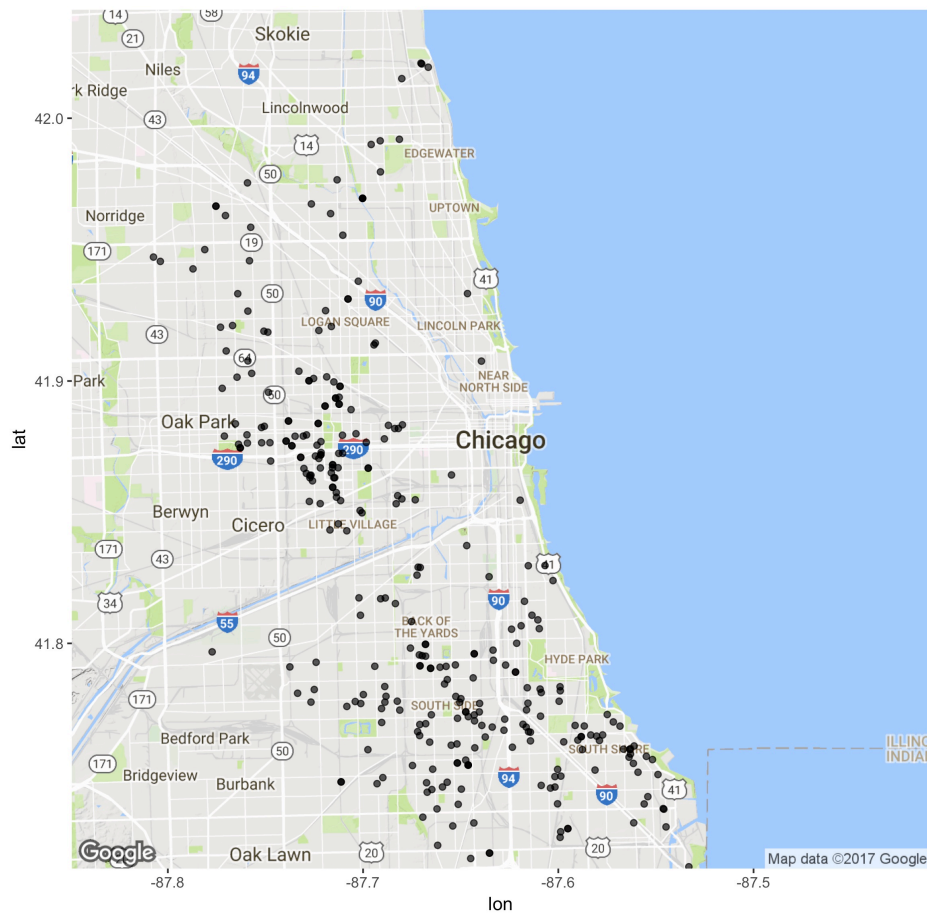


Figure 3: Area of CFO Referrals: Chicago

With this map, also possible to add layers of data. For example, we can pull crime data from the [City of Chicago](#) and overlay that information on the referrals to see how our young people are affected by violence in their communities.

This map shows the density of crime throughout Chicago. More lines indicate a higher density of crime. The city's data has a total of 31 categories of crime. Additional analysis would depend on further cleaning and refinement of this data. This work would allow help showcase density of violent crimes, property crimes, or drug-related crimes, each of which may have an influence on CFO attendance. Other layers of geographic information, such as CFO site locations, could be added to these maps to help identify more of the difficulties young people have in attending CFO programming.

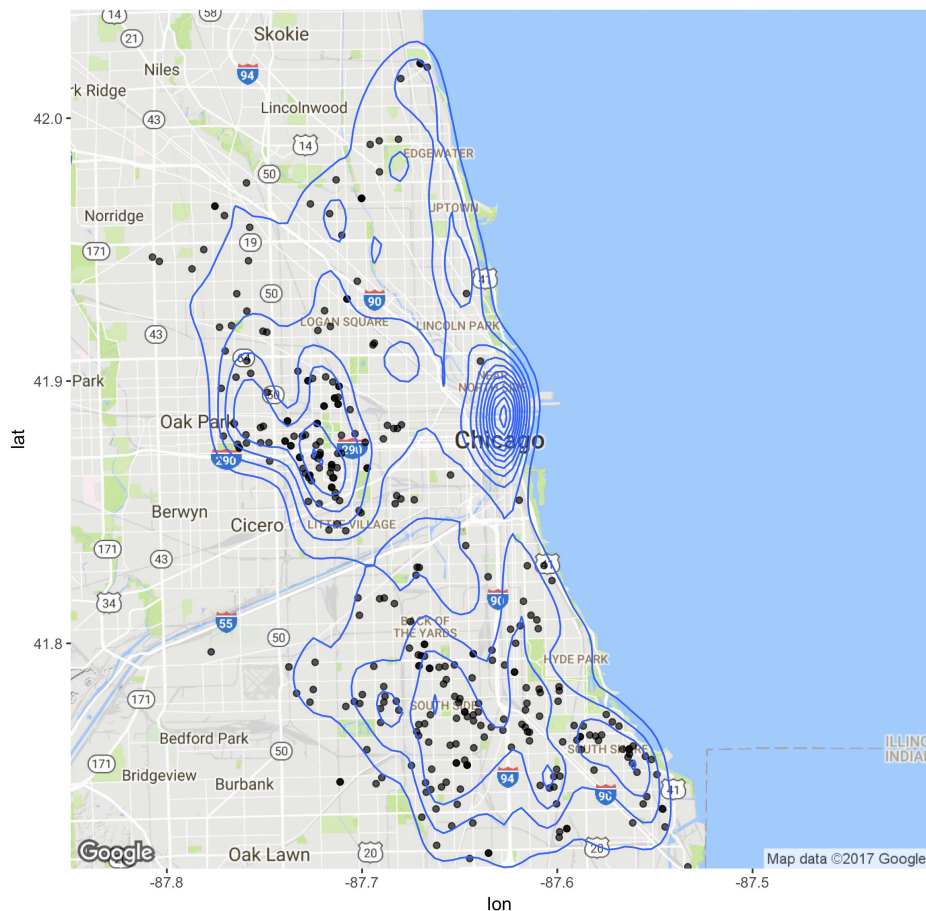


Figure 4: Crime Countour and CFO Referrals

Summary and Solutions

Improving CFO attendance is critical to PYD, and the key to improving attendance is understanding the factors that impact each site and CFO. Analysis of the data indicates that while a majority of young people invited to CFOs do not attend the program, the proportions vary greatly between units and CFO locations. The attendance rates, when broken down by Zipcode and PO, suggests levels of buy-in that can be leveraged for improved attendance as well as future CFO programming. With this goal in mind, the following domain-specific suggestions should be implemented. These are bottom-lined as follows:

JEMS

1. PYD should get query-level access to JEMS
 - Failing that, the exported spreadsheet should include:
 - Dates assigned to CFOs
 - Dates a young person attended a CFO
 - Exporting the JEMSID search results to determine what kind of impact CFOs may, or may not, have had
2. JEMS should be able to sort by Zipcode

3. JEMS should accept multiple dates for multiple CFOs per month
 - As of this writing, it is difficult complete attendance in JEMS, given the variety of dates, locations, and the drawbacks of associating a CFO date with

Tracking and Storing Data

JEMS and Excel are the standard tools for:

- Assigning youth to a CFO
 - Excel is critical for grouping by zipcode
 - Mail merging the letters mailed to youth and POs
- Attendance tracking
 - Attendance is hand written on printed spreadsheets
 - These notes are inputted into JEMS
 - Paper files are put in binders for indefinite storage

To improve analysis, either JEMS needs to be modified to take additional attendance conditions (Yes, No, In Custody/Adult Custody, Safety Issues, Out of State, Not Appropriate for CFO) and export them to Excel, or the following numbers should be used in the excel spread sheets:

1. Yes
2. No
3. In Custody
4. In Adult Custody
5. No due to Safety Conditions
6. Out of State
7. Not Appropriate for CFO

Additionally, for items 5-7, JEMS should be modified so those young people with those codes will not be required to attend additional CFOs.

Currently the department stores CFO records in Excel files, JEMS, and in binders. If JEMS is not updated to incorporate the additional attendance records, the department should considering other digital storage methods (specifically some sort of database, including Access or some version of SQL.)

Reporting

While Excel and MS Word are the go-to reporting tools for this kind of work in the department currently, this report also showcases the ability of *R* to import, clean, and transform raw data into reportable information. Map creation in *R* is also superior to the ability of MS Excel. The formatting drawbacks of *R* can be mitigated by experience and by creating department specific reporting templates. This goes beyond the scope of this report, but if the department was interested in using these free tools, significant resources are available for support and training.

Bottomline

CFOs are the future of youth led programming. To properly expand and assess CFOs, additional data analysis tools are required. Since April, 2017:

- 122 young people were invited to attend CFOs
- CFOs covered the following zipcodes 60827, 60617, 60621, 60628, 60636, 60619, 60649, 60651, 60644
- The attendance ratio is a 25.4%, with a significant degree of variance between POs and Zipcodes.

Furthermore, MS Excel is currently necessary for notifying clients and POs about the CFO date; however, analysis should be done in the *R* computer language. This will cut down on the time it takes to generate reports, as both the statistics and the reports can generated straight from *R*. This language also allows for fast and easy graphs, including maps of crime data. Where as CFOs are the future of youth led reporting, *R* should be the future of research division of the department.

Table 6: Attendance of Youth, Per Zip and Probation Officer

POLNAME	Attended	Absent	Invited	60617	60619	60621	60628	60636	60644	60649	60651
Akis	3	14	17	NA	NA	NA	NA	17.6%	NA	NA	NA
Akis	4	19	23	NA	NA	17.4%	NA	NA	NA	NA	NA
Almaraz	3	14	17	NA	NA	NA	NA	17.6%	NA	NA	NA
Almaraz	4	19	23	NA	NA	17.4%	NA	NA	NA	NA	NA
Bailey	4	13	17	NA	NA	NA	23.5%	NA	NA	NA	NA
Bloechel	3	12	15	NA	NA	NA	NA	NA	NA	20.0%	NA
Bloechel	7	5	12	58.3%	NA	NA	NA	NA	NA	NA	NA
Bloechel	7	11	18	NA	38.9%	NA	NA	NA	NA	NA	NA
Bohl	1	11	12	NA	NA	NA	NA	NA	NA	NA	8.3%
Bohl	2	11	13	NA	NA	NA	NA	NA	15.4%	NA	NA
Bowe	3	14	17	NA	NA	NA	NA	17.6%	NA	NA	NA
Bowe	4	13	17	NA	NA	NA	23.5%	NA	NA	NA	NA
Bowe	4	19	23	NA	NA	17.4%	NA	NA	NA	NA	NA
Bowling	3	12	15	NA	NA	NA	NA	NA	NA	20.0%	NA
Bowling	7	11	18	NA	38.9%	NA	NA	NA	NA	NA	NA
Brown	4	19	23	NA	NA	17.4%	NA	NA	NA	NA	NA
Butler	3	12	15	NA	NA	NA	NA	NA	NA	20.0%	NA
Butler	7	11	18	NA	38.9%	NA	NA	NA	NA	NA	NA
Campbell	1	11	12	NA	NA	NA	NA	NA	NA	NA	8.3%
Carroll	4	13	17	NA	NA	NA	23.5%	NA	NA	NA	NA
Casey	1	11	12	NA	NA	NA	NA	NA	NA	NA	8.3%
Coley	2	11	13	NA	NA	NA	NA	NA	15.4%	NA	NA
Cox	4	13	17	NA	NA	NA	23.5%	NA	NA	NA	NA
Dawkins	3	14	17	NA	NA	NA	NA	17.6%	NA	NA	NA
Dawkins	4	19	23	NA	NA	17.4%	NA	NA	NA	NA	NA
Gaston	7	11	18	NA	38.9%	NA	NA	NA	NA	NA	NA
Gonzales	4	13	17	NA	NA	NA	23.5%	NA	NA	NA	NA
Herner	4	13	17	NA	NA	NA	23.5%	NA	NA	NA	NA
Hopkins	2	11	13	NA	NA	NA	NA	NA	15.4%	NA	NA
Hudson	1	11	12	NA	NA	NA	NA	NA	NA	NA	8.3%
Jones	3	12	15	NA	NA	NA	NA	NA	NA	20.0%	NA
Jones	3	14	17	NA	NA	NA	NA	17.6%	NA	NA	NA
Jones	7	11	18	NA	38.9%	NA	NA	NA	NA	NA	NA
Junious	4	13	17	NA	NA	NA	23.5%	NA	NA	NA	NA
Korte	4	19	23	NA	NA	17.4%	NA	NA	NA	NA	NA
Krzeminski	4	13	17	NA	NA	NA	23.5%	NA	NA	NA	NA
Lacoy	2	11	13	NA	NA	NA	NA	NA	15.4%	NA	NA
Lewis	7	11	18	NA	38.9%	NA	NA	NA	NA	NA	NA
Morton	3	12	15	NA	NA	NA	NA	NA	NA	20.0%	NA
Morton	7	11	18	NA	38.9%	NA	NA	NA	NA	NA	NA
Muhammad	4	19	23	NA	NA	17.4%	NA	NA	NA	NA	NA
Muhammad	7	11	18	NA	38.9%	NA	NA	NA	NA	NA	NA
Naegele	4	13	17	NA	NA	NA	23.5%	NA	NA	NA	NA
Newton-Hart	1	11	12	NA	NA	NA	NA	NA	NA	NA	8.3%
Nieto	1	11	12	NA	NA	NA	NA	NA	NA	NA	8.3%
Ollins	1	11	12	NA	NA	NA	NA	NA	NA	NA	8.3%
Perfetti	7	5	12	58.3%	NA	NA	NA	NA	NA	NA	NA
Pulido	1	11	12	NA	NA	NA	NA	NA	NA	NA	8.3%
Purdy	3	12	15	NA	NA	NA	NA	NA	NA	20.0%	NA
Purdy	4	19	23	NA	NA	17.4%	NA	NA	NA	NA	NA
Purdy-Blake	3	12	15	NA	NA	NA	NA	NA	NA	20.0%	NA
Rivera	3	12	15	NA	NA	NA	NA	NA	NA	20.0%	NA
Rivera	7	5	12	58.3%	NA	NA	NA	NA	NA	NA	NA
Roberts	3	12	15	NA	NA	NA	NA	NA	NA	20.0%	NA
Rodriguez	7	11	18	NA	38.9%	NA	NA	NA	NA	NA	NA
Scalzetti	3	14	17	NA	NA	NA	NA	17.6%	NA	NA	NA
Scalzetti	4	19	23	NA	NA	17.4%	NA	NA	NA	NA	NA
Siegel	3	14	17	NA	NA	NA	NA	17.6%	NA	NA	NA
Siegel	4	19	23	NA	NA	17.4%	NA	NA	NA	NA	NA