

Data, Research, and Visualization Unit

Executive Summary

The department should invest staffing resources into a Data, Research, and Visualization unit that, in addition to compiling AOIC statistics, works with the [Cook County Geographic Information Systems Department](#) to create new, data driven tools and supplement the dashboards of created by [CFive's Supervisor](#). This unit, staffed by sworn staff and support professionals would use Excel, [R](#), and [GIS](#) to collect, clean, and analyze data to ensure the department is maximizing the investment in *CFive Supervisor*.

The Issue

To demonstrate the need of a new Data, Research, and Visualization unit, consider the current report for AOIC. Currently, the AOIC Stats are presented in the following format:

Table 1: First Eight Columns of CFive Supervisor

X__1	X__2	X__3	X__4	X__5	X__6	X__7	X__8
Terri Griffin	NA	NA	NA	NA	NA	NA	NA
Police District 17, 14, 18-20 - Cal. 58, 60	NA	NA	NA	NA	NA	NA	NA
Diane Bufano	NA	NA	NA	NA	NA	NA	NA
NA	Laura Donnelly	Andrea Korte	Michelle Malave	Amy O'Rourke	Yvonne Pulido	Diane Bufano	NA
Total Caseload	17	9	20	13	21	1	NA
YASI HIGH	4	2	3	2	5	NA	NA

Of course, the Excel report is more aesthetically pleasing, but the overall format is identical: Multiple units on the document, multiple breakdowns of stats collected, and multiple types of observations.

The first eight columns are the POs, per unit, at a glance, depending on the number of POs in the unit and the number of cases assigned to the Supervisor.

Table 2: Caseload Aggregate Data

X__9	X__10	X__11
NA	NA	NA
NA	NA	NA
NA	NA	NA
CASELOAD TOTAL	AVERAGE CASELOAD	Total P.Os
81	16	5
16	NA	NA

Columns 9 through 10 are aggregates of the previous unit data.

Table 3: VOPs and Caseload Averages

X_12	X_13	VOP'S FILED BY UNIT	Baseline(Average for the month)	Caseload size	Dept average caseload size	Average Caseload Size for Unit	Average Caseload Size
1	SPO Bufano	2	2	81	72	16	16
2	SPO Carter	2	2	29	72	14.5	16
3	SPO Flanagan	2	2	59	72	14	16
4	SPO Alejo	3	2	58	72	11.6	16
5	SPO Herner	1	2	46	72	15.333333333333334	16
6	SPO Moore	4	2	55	72	11	16

The remaining seven columns are a mishmash of stats by unit and department.

Presenting AOIC data in this fashion limits further investigation of data. Multiple spreadsheets must be created to combine data for month-month comparisons, and compiling said spreadsheets increases the chances of errors due to spelling mistakes, accidentally deleting spreadsheet structure, or over-writing spreadsheets while saving them.

Fortunately, [CFive's Supervisor](#) will be able to solve most of these data collection issues. As of the current understanding of the Statement of Work, *Supervisor* will pull these necessary numbers from the system as a standard report. Another possible scenario is SPOs will complete a *Supervisor* based form that will guide the creation of the report. Regardless, the method in which the AOIC Stats are gathered and presented will change. Unfortunately, the reporting features are not completed and not available for preview.

This gap presents an opportunity for the department to reformat any and all of the statistical forms and reports, starting with the AOIC reports, to ensure our data and visualization needs are not just met, but exceeded. The *Supervisor's Go-Live* ¹ date provides an incentive to get this material completed quickly.

New Format and Methods: The Solution

Implementing [tidy data](#) is the first step to any transformation to a data-driven office. Tidy data is short hand for:

- Columns are single variable
- Rows are single observation
- Each observational unit forms a table

Currently, the AOIC spreadsheets violates all of the rules of tidy data. The easiest example to understand is how each sheet is a collection of observations and aggregates across units. In truth, each geographic unit should have their own worksheet, with separate worksheets for averages and other calculated variables. It appears that the sheets follow the first and second rules; however, considering what is being tracked, the columns and rows are transposed. Put differently, in the current format, POs are variables when they they are actually tracking observations. In tidy terms this makes them *rows*.

¹As of this writing, Go-Live is set for March 29, 2019

GBO As a Demo

For example, take the GBO unit. On the standard AOIC spreadsheet, GBO's stats are between rows 401 and 418:

Table 4: September, 2018

AOIC_Stat	PO_Stat1	PO_Stat2	PO_Stat3	PO_Stat4	PO_Stat5	PO_Stat6	PO_Stat7
Melissa Spooner	NA	NA	NA	NA	NA	NA	NA
Police district 2 - Cal 55	NA	NA	NA	NA	NA	NA	NA
Lloyd Marshall	NA	NA	NA	NA	NA	NA	NA
NA	Ernest Jones	Michael Muhammad	Rodney Purdy-Blake	James Smith	ADMIN	NA	NA
Total Caseload	16	17	15	16	NA	NA	NA
YASI HIGH	7	8	4	1	NA	NA	NA
YASI MOD	6	6	7	9	NA	NA	NA
YASI LOW	1	2	2	3	NA	NA	NA
YASI UNCLASSIFIED	2	1	2	3	NA	NA	NA
Scheduled Termination	2	1	1	1	NA	NA	NA
Early Termination	NA	NA	NA	NA	NA	NA	NA
Revoked-Technical/DOC-IDJJ	NA	NA	NA	NA	NA	NA	NA
Revoked-new offense/DOC-IDJJ	NA	NA	NA	NA	NA	NA	NA
Unsatisfactory termination	NA	NA	NA	NA	NA	NA	NA
Transferred out	NA	NA	NA	NA	NA	NA	NA
Social investigation	NA	NA	NA	NA	NA	NA	NA
Supplemental social	NA	NA	NA	NA	NA	NA	NA
VOP Filed	NA	NA	NA	NA	NA	NA	NA

At a glance, one can see how individual officers case loads have changed over the previous month. However, if one was to compare GBO and Lawndale(rows 201-218), the issue with messy data becomes apparent:

GBO and Lawndale

Table 5: GBO and Lawndale: September, 2018

AOIC_Stat	PO_Stat1	PO_Stat2	PO_Stat3	PO_Stat4	PO_Stat5	PO_Stat6	PO_Stat7
Melissa Spooner	NA	NA	NA	NA	NA	NA	NA
Police district 2 - Cal 55	NA	NA	NA	NA	NA	NA	NA
Lloyd Marshall	NA	NA	NA	NA	NA	NA	NA
NA	Ernest Jones	Michael Muhammad	Rodney Purdy-Blake	James Smith	ADMIN	NA	NA
Total Caseload	16	17	15	16	NA	NA	NA
YASI HIGH	7	8	4	1	NA	NA	NA
YASI MOD	6	6	7	9	NA	NA	NA
YASI LOW	1	2	2	3	NA	NA	NA
YASI UNCLASSIFIED	2	1	2	3	NA	NA	NA
Scheduled Termination	2	1	1	1	NA	NA	NA
Early Termination	NA	NA	NA	NA	NA	NA	NA
Revoked-Technical/DOC-IDJJ	NA	NA	NA	NA	NA	NA	NA
Revoked-new offense/DOC-IDJJ	NA	NA	NA	NA	NA	NA	NA
Unsatisfactory termination	NA	NA	NA	NA	NA	NA	NA
Transferred out	NA	NA	NA	NA	NA	NA	NA
Social investigation	NA	NA	NA	NA	NA	NA	NA
Supplemental social	NA	NA	NA	NA	NA	NA	NA
VOP Filed	NA	NA	NA	NA	NA	NA	NA
Jose Isais	NA	NA	NA	NA	NA	NA	NA
Police Distret 11 - Cal. 57	NA	NA	NA	NA	NA	NA	NA
Aaron Campbell	NA	NA	NA	NA	NA	NA	NA
NA	Laterrian Hill	Rance Hopkins*	Pamela Hudson	Tesa Newton-Hart	Kenneth Ollins	Campbell	NA
Total Caseload	27	31	23	22	26	1	NA
YASI HIGH	9	NA	6	10	3	NA	NA
YASI MOD	10	NA	9	5	11	1	NA
YASI LOW	8	NA	3	6	9	NA	NA
YASI UNCLASSIFIED	NA	31	5	1	3	NA	NA
Scheduled Termination	2	NA	1	2	NA	NA	NA
Early Termination	NA	NA	NA	NA	NA	NA	NA
Revoked-Technical/DOC-IDJJ	NA	NA	NA	NA	NA	NA	NA
Revoked-new offense/DOC-IDJJ	NA	NA	NA	NA	NA	NA	NA
Unsatisfactory termination	2	NA	NA	NA	1	NA	NA
Transferred out	NA	9	NA	NA	NA	NA	NA
Social investigation	NA	8	NA	NA	NA	NA	NA
Supplemental social	NA	NA	NA	NA	NA	NA	NA
VOP Filed	1	NA	NA	NA	NA	1	NA

With only one additional unit, it is relatively easy to compare and contrast intra-unit data. It does require a reader to track data vertically and horizontally. It is far from ideal. Specifically, one has to read horizontally and vertically to compare baseline data month over month. With each unit added to this format, the more difficult it becomes to not just compare data points, but to find data points. For example:

Table 6: GBO, Lawndale, and 8th District: September, 2018

AOIC_Stat	PO_Stat1	PO_Stat2	PO_Stat3	PO_Stat4	PO_Stat5	PO_Stat6	PO_Stat7
Melissa Spooner	NA	NA	NA	NA	NA	NA	NA
Police district 2 - Cal 55	NA	NA	NA	NA	NA	NA	NA
Lloyd Marshall	NA	NA	NA	NA	NA	NA	NA
NA	Ernest Jones	Michael Muhammad	Rodney Purdy-Blake	James Smith	ADMIN	NA	NA
Total Caseload	16	17	15	16	NA	NA	NA
YASI HIGH	7	8	4	1	NA	NA	NA
YASI MOD	6	6	7	9	NA	NA	NA
YASI LOW	1	2	2	3	NA	NA	NA
YASI UNCLASSIFIED	2	1	2	3	NA	NA	NA
Scheduled Termination	2	1	1	1	NA	NA	NA
Early Termination	NA	NA	NA	NA	NA	NA	NA
Revoked-Technical/DOC-IDJJ	NA	NA	NA	NA	NA	NA	NA
Revoked-new offense/DOC-IDJJ	NA	NA	NA	NA	NA	NA	NA
Unsatisfactory termination	NA	NA	NA	NA	NA	NA	NA
Transferred out	NA	NA	NA	NA	NA	NA	NA
Social investigation	NA	NA	NA	NA	NA	NA	NA
Supplemental social	NA	NA	NA	NA	NA	NA	NA
VOP Filed	NA	NA	NA	NA	NA	NA	NA
Jose Isais	NA	NA	NA	NA	NA	NA	NA
Police Distrc 11 - Cal. 57	NA	NA	NA	NA	NA	NA	NA
Aaron Campbell	NA	NA	NA	NA	NA	NA	NA
NA	Laterrian Hill	Rance Hopkins*	Pamela Hudson	Tesa Newton-Hart	Kenneth Ollins	Campbell	NA
Total Caseload	27	31	23	22	26	1	NA
YASI HIGH	9	NA	6	10	3	NA	NA
YASI MOD	10	NA	9	5	11	1	NA
YASI LOW	8	NA	3	6	9	NA	NA
YASI UNCLASSIFIED	NA	31	5	1	3	NA	NA
Scheduled Termination	2	NA	1	2	NA	NA	NA
Early Termination	NA	NA	NA	NA	NA	NA	NA
Revoked-Technical/DOC-IDJJ	NA	NA	NA	NA	NA	NA	NA
Revoked-new offense/DOC-IDJJ	NA	NA	NA	NA	NA	NA	NA
Unsatisfactory termination	2	NA	NA	NA	1	NA	NA
Transferred out	NA	9	NA	NA	NA	NA	NA
Social investigation	NA	8	NA	NA	NA	NA	NA
Supplemental social	NA	NA	NA	NA	NA	NA	NA
VOP Filed	1	NA	NA	NA	NA	1	NA
Ore Jones, III	NA	NA	NA	NA	NA	NA	NA
PolicePolice District 10 - Cal. 53	NA	NA	NA	NA	NA	NA	NA
Bennie Blair	NA	NA	NA	NA	NA	NA	NA
NA	Cedric Bell	Tyrone Hutson	Tandra Tyler	Robert Hillyer	Bennie Blair	NA	NA
Total Caseload	21	21	19	11	2	NA	NA
YASI HIGH	NA	5	2	4	2	NA	NA
YASI MOD	11	4	10	5	NA	NA	NA
YASI LOW	2	7	4	2	NA	NA	NA
YASI UNCLASSIFIED	8	5	3	NA	NA	NA	NA
Scheduled Termination	1	2	1	NA	NA	NA	NA
Early Termination	NA	NA	NA	NA	NA	NA	NA
Revoked-Technical/DOC-IDJJ	NA	NA	NA	NA	NA	NA	NA
Revoked-new offense/DOC-IDJJ	1	NA	NA	NA	NA	NA	NA
Unsatisfactory termination	1	NA	NA	NA	NA	NA	NA
Transferred out	1	1	3	1	NA	NA	NA
Social investigation	NA	NA	1	NA	NA	NA	NA
Supplemental social	NA	2	2	1	NA	NA	NA
VOP Filed	NA	NA	NA	1	NA	NA	NA

The current excel format solves this issue by adding two or three empty rows; however, the lack of standard difference between rows makes machine based analysis slower and more time consuming. It also makes plotting data exceptionally difficult.

New format

The solution looks like this:

Table 7: GBO: September, 2018

PO_Name	Early Termination	Revoked-new offense/DOC-IDJJ	Revoked-Technical/DOC-IDJJ	Scheduled Termination	Social investigation	Supplemental social	Transferred out	Unsatisfactory termination	VOP Filed	YASI HIGH	YASI LOW	YASI MOD	YASI UNCLASSIFIED	Unit	SPO	DCPO	
ADMIN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Police district 2 - Cal 55	Lloyd Marshall	Melissa Spooner
Ernest Jones	0	0	0	2	0	0	0	0	0	7	1	6	2	Police district 2 - Cal 55	Lloyd Marshall	Melissa Spooner	
James Smith	0	0	0	1	0	0	0	0	0	1	3	9	3	Police district 2 - Cal 55	Lloyd Marshall	Melissa Spooner	
Michael Muhammad	0	0	0	1	0	0	0	0	0	8	2	6	1	Police district 2 - Cal 55	Lloyd Marshall	Melissa Spooner	
Rodney Purdy-Blake	0	0	0	1	0	0	0	0	0	4	2	7	2	Police district 2 - Cal 55	Lloyd Marshall	Melissa Spooner	

This create a difficult-to-read 17 columns² wide table which is difficult to read on paper. The difficulty is mitigated because of the ability to to zoom in on particular areas of interest, either by automating the reports to fit particular needs or through web based technologies (which will be discussed later.) It is also possible to filter data by specific variables. For example, if one just wanted YASI data:

²The columns are: *PO_Name*, *Early Termination*, *Revoked-new offense/DOC-IDJJ*, *Revoked-Technical/DOC-IDJJ*, *Scheduled Termination*, *Social investigation*, *Supplemental social*, *Transferred out*, *Unsatisfactory termination*, *VOP Filed*, *YASI HIGH*, *YASI LOW*, *YASI MOD*, *YASI UNCLASSIFIED*, *Unit*, *SPO*, *DCPO*

Table 8: Case Totals, GBO: September, 2018

PO_Name	High	Mod	Low	Unclassified	Total
ADMIN	0	0	0	0	0
Ernest Jones	7	6	1	2	16
James Smith	1	9	3	3	16
Michael Muhammad	8	6	2	1	17
Rodney Purdy-Blake	4	7	2	2	15

7

Once the Excel data is in this format, it is easier to analyse. For instance, take a table of two units arranged by High risk cases:

Table 9: GBO and Lawndale by High Risk: September, 2018

PO_Name	Early Termination	Revoked-new offense/DOC-IDJJ	Revoked-Technical/DOC-IDJJ	Scheduled Termination	Social investigation	Supplemental social	Transferred out	Unsatisfactory termination	VOP Filed	YASI HIGH	YASI LOW	YASI MOD	YASI UNCLASSIFIED	Unit	SPO	DCPO
Tessa Newton-Hart	0	0	0	2	0	0	0	0	0	10	6	5		1 Police District 11 - Cal. 57	Aaron Campbell	Jose Isais
Laterrian Hill	0	0	0	2	0	0	0	0	2	1	9	8	10	0 Police District 11 - Cal. 57	Aaron Campbell	Jose Isais
Michael Muhammad	0	0	0	1	0	0	0	0	0	8	2	6		1 Police district 2 - Cal 55	Lloyd Marshall	Melissa Spooner
Ernest Jones	0	0	0	2	0	0	0	0	0	7	1	6		2 Police district 2 - Cal 55	Lloyd Marshall	Melissa Spooner
Pamela Hudson	0	0	0	1	0	0	0	0	0	6	3	9		5 Police District 11 - Cal. 57	Aaron Campbell	Jose Isais
Rodney Purdy-Blake	0	0	0	1	0	0	0	0	0	4	2	7		2 Police district 2 - Cal 55	Lloyd Marshall	Melissa Spooner
Kenneth Ollins	0	0	0	0	0	0	0	0	1	0	3	9	11	3 Police District 11 - Cal. 57	Aaron Campbell	Jose Isais
James Smith	0	0	0	1	0	0	0	0	0	1	3	9		3 Police district 2 - Cal 55	Lloyd Marshall	Melissa Spooner
Campbell	0	0	0	0	0	0	0	0	0	1	0	0	1	0 Police District 11 - Cal. 57	Aaron Campbell	Jose Isais
Rance Hopkins*	0	0	0	0	8	0	9	0	0	0	0	0		31 Police District 11 - Cal. 57	Aaron Campbell	Jose Isais
ADMIN	0	0	0	0	0	0	0	0	0	0	0	0		0 Police district 2 - Cal 55	Lloyd Marshall	Melissa Spooner

Crowding all of the columns onto one page is not ideal; however, one can easily see all of the stats by unit for a given month. Filtering by case load stats becomes exceptionally easier:³

Table 10: Case Totals, Lawndale and GBO: September, 2018

PO_Name	Unit	High	Mod	Low	Unclassified	Total
Rance Hopkins*	Polce Distrct 11 - Cal. 57	0	0	0	31	31
Laterrian Hill	Polce Distrct 11 - Cal. 57	9	10	8	0	27
Kenneth Ollins	Polce Distrct 11 - Cal. 57	3	11	9	3	26
Pamela Hudson	Polce Distrct 11 - Cal. 57	6	9	3	5	23
Tesa Newton-Hart	Polce Distrct 11 - Cal. 57	10	5	6	1	22
Michael Muhammad	Police district 2 - Cal 55	8	6	2	1	17
Ernest Jones	Police district 2 - Cal 55	7	6	1	2	16
James Smith	Police district 2 - Cal 55	1	9	3	3	16
Rodney Purdy-Blake	Police district 2 - Cal 55	4	7	2	2	15
Campbell	Polce Distrct 11 - Cal. 57	0	1	0	0	1
ADMIN	Police district 2 - Cal 55	0	0	0	0	0

³Please note the spelling errors come from excel and not the code written in this document.

Plots and Graphs

Like Excel, this format can be plotted:

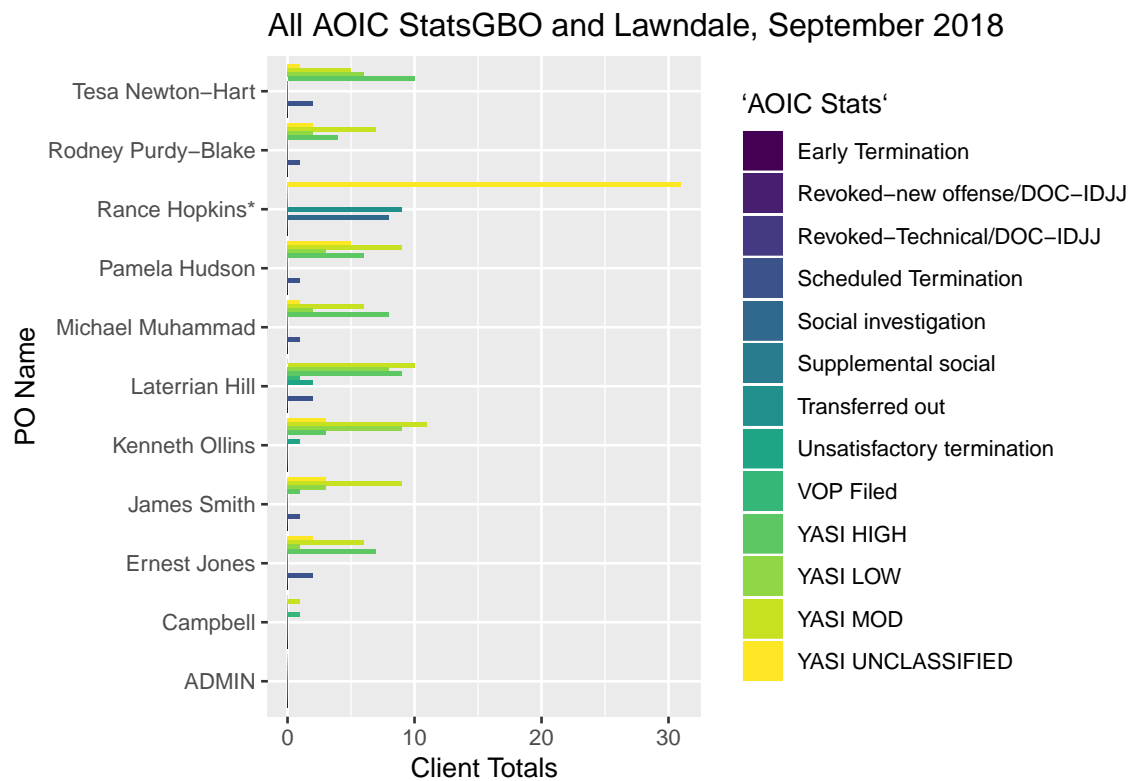


Figure 1: GBO and Lawndale, September 2018

And unlike Excel, this format easily lends itself to additional visualizations, such as filtering results to focus on specific variables, such as Social Investigations and VOP's Filed. For this example, the background code is included to show how this work is done.

First as a chart:

```
cal_57 %>%
  tidy_up() %>%
  bind_rows(lloyd) %>%
  select(1, `Social Investigation` = `Social investigation`,
         `VOP` = `VOP Filed`) %>%
  filter(`Social Investigation` > 0 | VOP > 0) %>%
  kable(booktabs = TRUE, format = "latex",
        caption = "GB0: September, 2018") %>%
  kable_styling(latex_options = c("striped",
                                   "hold_position", "scale_down"),
                stripe_color = "lightgray")
```

Table 11: GBO: September, 2018

PO_Name	Social Investigation	VOP
Campbell	0	1
Laterrian Hill	0	1
Rance Hopkins*	8	0

And as a graph:

```
lloyd <- cal_55_2 %>%
  tidy_up()

cal_57 %>%
  tidy_up() %>%
  bind_rows(lloyd) %>%
  select(1, `Social Investigation` = `Social investigation`,
         `VOP` = `VOP Filed`) %>%
  gather(c(`Social Investigation`, `VOP`), key = Social_VOP,
         value = Totals) %>%
  ggplot(aes(x = PO_Name, y = Totals, fill = `Social_VOP`)) +
  geom_bar(stat = "identity", position = "dodge") +
  viridis::scale_fill_viridis(discrete = TRUE, name = "Socials and VOP") +
  labs(title = "GB0 and Lawndale, September 2018",
```

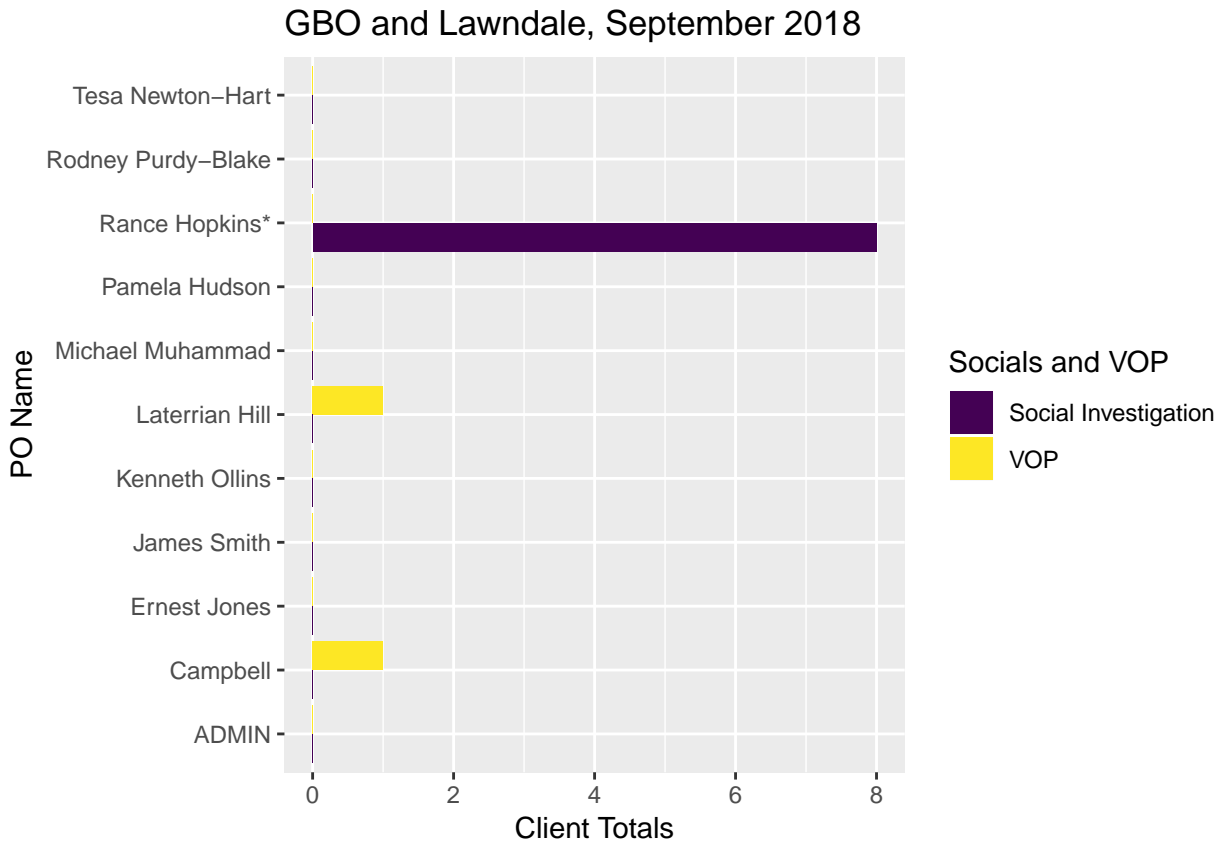


Figure 2: Socials and VOPs in GBO and Lawndale

```
x = "PO Name", y = "Client Totals") +
coord_flip()
```

In both of the previous examples, snippets of code were included in the document to showcase how R handles filtering. What makes this method superior to Excel is that the above code can be reused with a similar dataset. For instance, plotting the same AOIC stas with Markham North:

```
cal_mark_n %>%
  tidy_up() %>%
  select(1, `Social Investigation` = `Social investigation`,
         `VOP` = `VOP Filed`) %>%
  gather( c(`Social Investigation`, `VOP`), key = Social_VOP,
         value = Totals) %>%
  ggplot(aes(x = PO_Name, y = Totals, fill = `Social_VOP`)) +
  geom_bar(stat = "identity", position = "dodge") +
  viridis::scale_fill_viridis(discrete = TRUE, name = "Socials and VOP") +
  labs(title = "Markham North, September 2018",
       x = "PO Name", y = "Client Totals") +
  coord_flip()
```

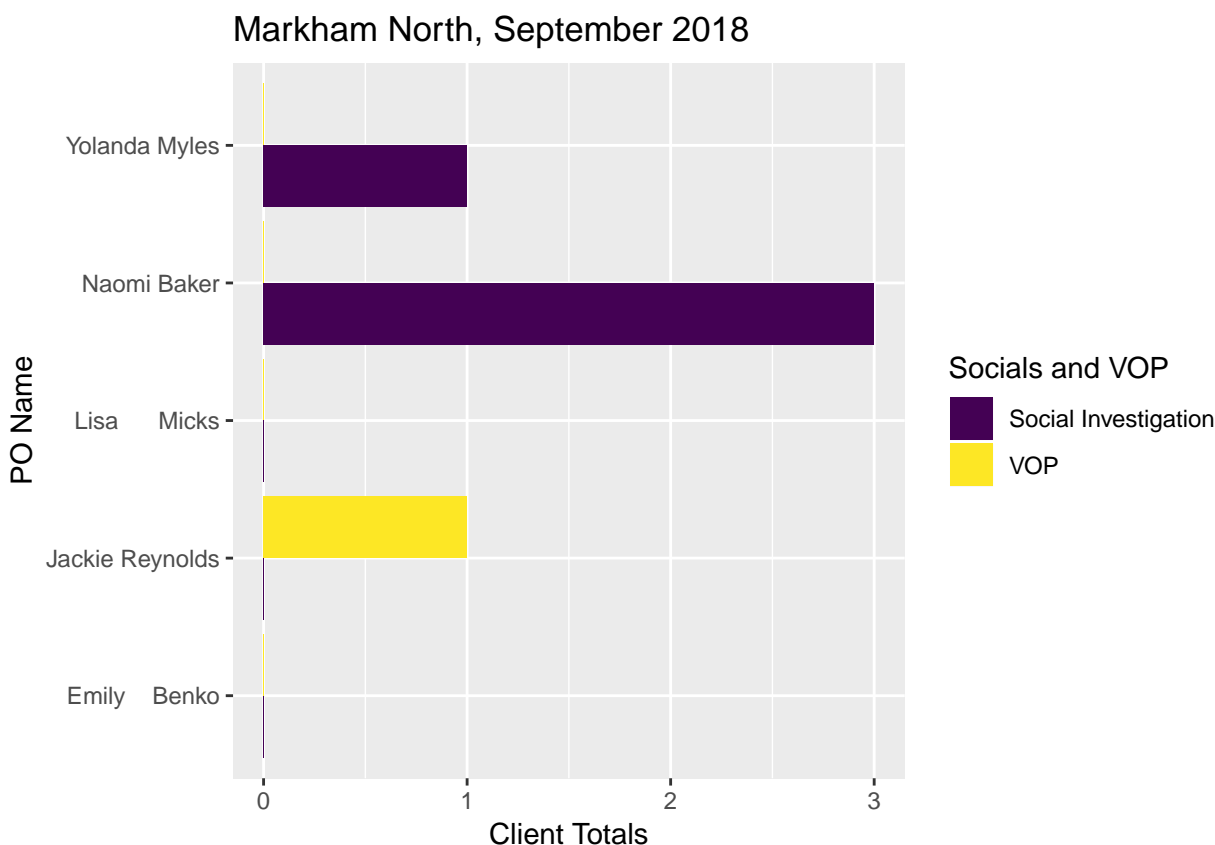


Figure 3: Socials and VOPs in Markham North

With Excel, each month's needs to be recalculated from the specific sheet. With R, the code can be reused with limited changes.

With additional resources, specifically installing free software on the Chief Judge's server, the department would be able to run these same reports internally as a website. This software, known as [Shiny](#), would allow us to create dashboards and programmable reports that can be updated quickly. Additional examples of how Shiny can be used can be found [here](#).

Resources

CFive Supervisor will encourage the department to continue down a data driven path. Maximizing the return on this investment requires additional departmental resources, the minimum of which is a system administrator to ensure data integrity. Just doing the minimum, however, will not make the most of this new system (and the changes to procedure that is sure to follow.) To this end, a new unit should be created to ensure that handles all department data and visualizations. This would include the short term goals of changing the AOIC stat format to a Tidy one and as well as producing stats until CFive Supervisor's

is fully implemented, as well as long term goals of using our data to ensure the creation of data driven policies, procedures, and programs:

It is the policy of the Cook County Juvenile Probation and Court Services Department to maintain the highest level of data integrity within the department and with community partners and to use this data to develop policies, procedures, and programs that enhance the possibility of long term success for all young people under the care of the court.

This unit would house:

- Two support professionals to help collect data and generate reports and visualizations.
- The CFive System Administrator

Two such individuals already do this work in the department, so including them in this new department would only service to solidify this commitment to data driven approaches. The role of CFive Systems Administrator will need to be a permanent position for maintaining the system; therefore it is logical to add this position to a unit that is focused on data.

In addition to reporting and collection tasks, the supervisor of the unit would also be responsible for maintain and leverage all of the data sharing agreements the department has with other justice and community agencies. Other staff maybe needed to help develop programs that stem from the insights gained by maximizing the utility of **CFive Supervisor**; shifting staff resources to this new unit would suffice for current needs. In creating this unit, the department frames data as a “mission critical” aspect, and redistributes all of the varied data tasks under one roof. It is not a revolutionary change, simple an evolution of the departments needs, capabilities, and opportunities.