# Stats and the Future of the Department

*Martin Gleason, MS*

*May 31, 2018*

## What is R

R is a popular, functional programming language that is used for data analysis and data science. It allows for fast, reliable, and repeatable calculations.

### R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

## Including CFO DATA

For this demonstration, June CFO data was included.The first ten young people referred this month are as follows:

Table 1: Raw June CFO Referrals

| FNAME | LNAME | STRTNO | STRTDC | STRTNM | APRTNO | CITY | STATE | ZIP | SEX | UPDATE_IDENT | JEMSID | POFNAME | POLNAME |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DEMARCUS | ABRAM | 13240 | S | CORLISS | NA | CHICAGO | IL | 60827 | M | 1 | 10326791 | JOSEPH | NAEGELE |
| LATRELL | ALLEN | 5616 | S | ABERDEEN ST | NA | CHICAGO | IL | 60621 | M | 1 | 10335541 | HANNAH | SIEGEL |
| XAVIER | ALLEN | 10727 | S | CALUMET AVE | UNIT 1 | CHICAGO | IL | 60628 | M | 1 | 10342391 | VICTOR | JUNIOUS |
| CAPRIA | ALLEN | 5622 | S | HONORE ST | HOUSE | CHICAGO | IL | 60636 | F | 1 | 10316032 | DARRYL | DAWKINS |
| JULIAN | ALMANZA | 2511 | N | LAWNDALE AVE | NA | CHICAGO | IL | 60647 | M | 1 | 10341183 | ANTWAN | JONES |
| JOSE | ALVAREZ | 3567 | W | PALMER STREET | D | CHICAGO | IL | 60647 | M | 1 | 10347745 | YVONNE | PULIDO |
| JAYLEN | ANDERSON | 103 | E | 58TH ST | 3W | CHICAGO | IL | 60637 | M | 1 | 10350803 | MICHAEL | MUHAMMAD |
| TIYANA | ASH | 4348 | W | WILCOX STREET | 1ST FLOOR | CHICAGO | IL | 60624 | F | 1 | 10348582 | KENNETH | OLLINS |
| JESUS | AYALA | 2653 | S | CHRISTIANA AVE | NA | CHICAGO | IL | 60623 | M | 1 | 10327168 | HECTOR | ESCALERA |
| DALVIN | BALDWIN | 6409 | S | CARPENTER | 2ND FLOOR | CHICAGO | IL | 60621 | M | 1 | 10348209 | MICHAEL | MUHAMMAD |

### Messy Data

Note that this data is messy: Address is seperated into STRTNO, STRTDC, STRTNM, all of the headings are in caps, and some of the columns have no value for most officer's needs. While we can understand the raw data in context, what the department needs is clear information: Data processed in such a way that it is useable for everyone from line staff to management. This means we need to clean the data in order to get to more useful information.

## Tidy Data

R can transform this data quickly and reliably. The end results of this can be quickly loaded in to this document

```r
kable(head(cfo_june_cleaned, caption = "A Tidier Set of June CFO Data", align = c, trim = T)) %>%
  kable_styling(latex_options = c("striped", "scale_down", "hold_position"))
```

| FName | LName | Address | AptNo | City | State | Zip | Sex | JEMSID | POFName | POLName |
|-------|-------|---------|-------|------|-------|-----|-----|--------|---------|---------|
| DEMARCUS | ABRAM | 13240 S CORLISS | NA | CHICAGO | IL | 60827 | M | 10326791 | JOSEPH | NAEGELE |
| LATRELL | ALLEN | 5616 S ABERDEEN ST | NA | CHICAGO | IL | 60621 | M | 10335541 | HANNAH | SIEGEL |
| XAVIER | ALLEN | 10727 S CALUMET AVE | UNIT 1 | CHICAGO | IL | 60628 | M | 10342391 | VICTOR | JUNIOUS |
| CAPRIA | ALLEN | 5622 S HONORE ST | HOUSE | CHICAGO | IL | 60636 | F | 10316032 | DARRYL | DAWKINS |
| JULIAN | ALMANZA | 2511 N LAWNDALE AVE | NA | CHICAGO | IL | 60647 | M | 10341183 | ANTWAN | JONES |
| JOSE | ALVAREZ | 3567 W PALMER STREET | D | CHICAGO | IL | 60647 | M | 10347745 | YVONNE | PULIDO |

There are two things about this image that should be explained. First, the code chunk is turned on to illustrate how the software works. Secondly, the following table iss a low-level version of cleaning the data. It means that each young person refered to a CFO in JUNE is only represented once and only one kind of datum exists per cell. More plainly, tidy data means the data is properly organized and makes sense to both machines and people.

This means we can do more interesting analysis of the data. For example, this the data can display referrals by zipcode as a graph or as a table.
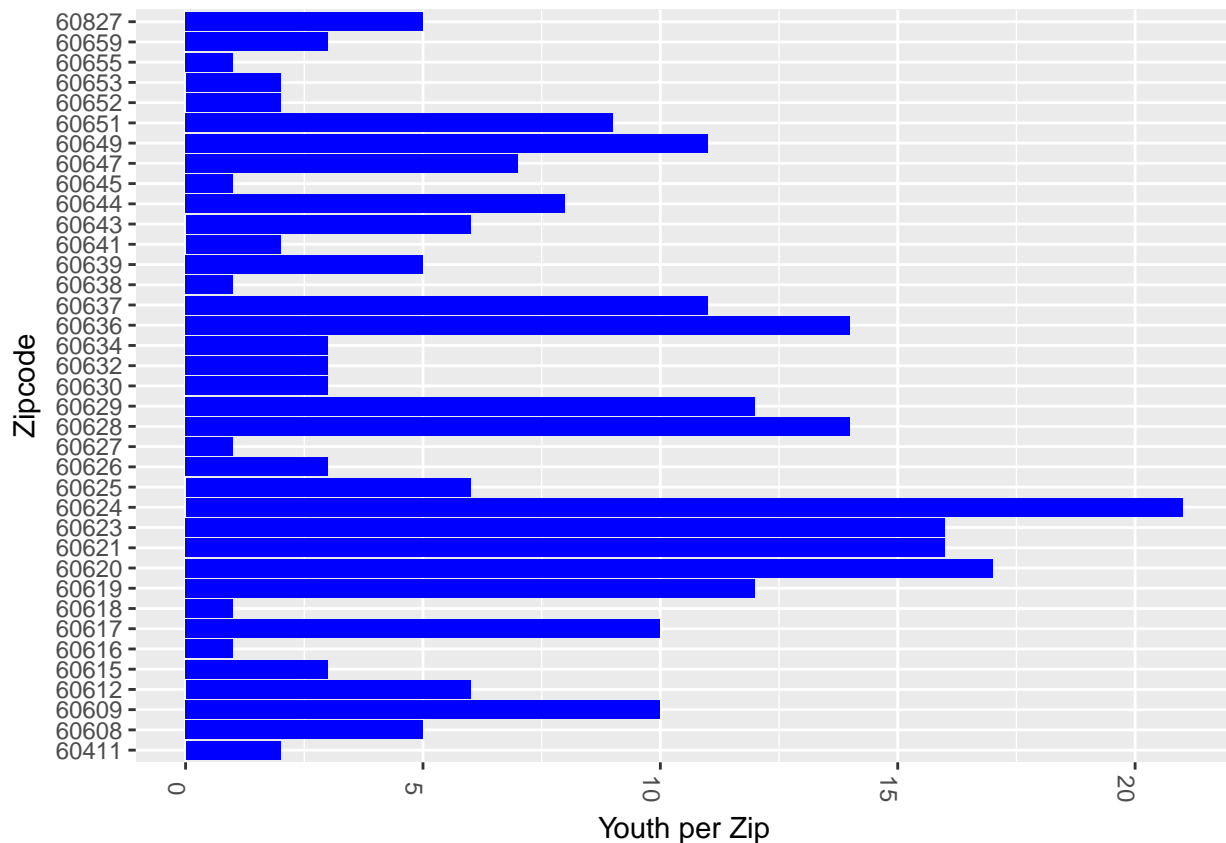


Figure 1: Referral by Zip

Table 2: June CFO: Referrals by Zip

| Zip | Youth_per_zip |
|---|---|
| 60624 | 21 |
| 60620 | 17 |
| 60621 | 16 |
| 60623 | 16 |
| 60628 | 14 |
| 60636 | 14 |
| 60619 | 12 |
| 60629 | 12 |
| 60637 | 11 |
| 60649 | 11 |
| 60609 | 10 |
| 60617 | 10 |
| 60651 | 9 |
| 60644 | 8 |
| 60647 | 7 |
| 60612 | 6 |
| 60625 | 6 |
| 60643 | 6 |
| 60608 | 5 |
| 60639 | 5 |
| 60827 | 5 |
| 60615 | 3 |
| 60626 | 3 |
| 60630 | 3 |
| 60632 | 3 |
| 60634 | 3 |
| 60659 | 3 |
| 60411 | 2 |
| 60641 | 2 |
| 60652 | 2 |
| 60653 | 2 |
| 60616 | 1 |
| 60618 | 1 |
| 60627 | 1 |
| 60638 | 1 |
| 60645 | 1 |
| 60655 | 1 |

All of this information is generated from the same software that crunches the numbers. Formatting these reports will take time; however, once they are formatted, adding and subtracting information will be significantly easier.
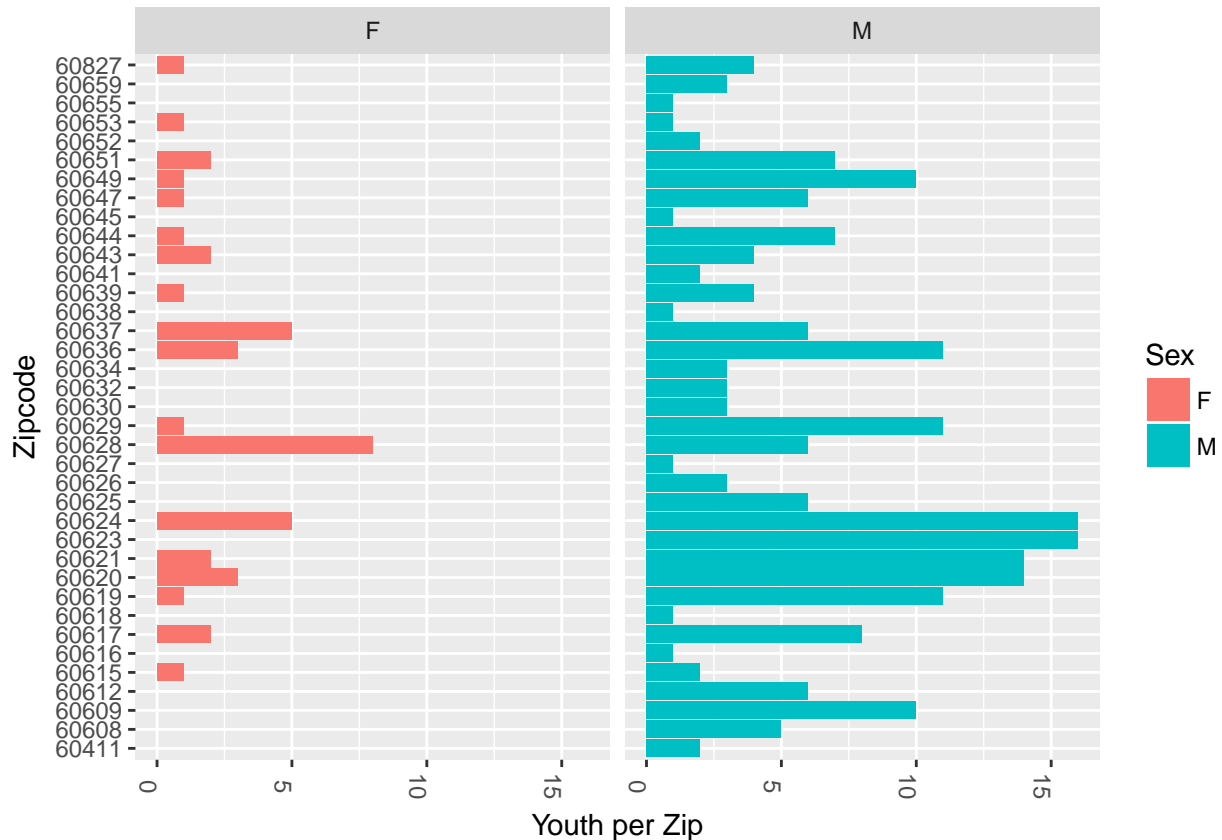
For example, we can also see which POs referred clients in June.

Table 3: Number of Referrals per PO From JEMS

| POFName | POLName | PO_Referrals |
|---|---|---|
| LISA | BLOECHEL | 10 |
| HANNAH | SIEGEL | 7 |
| RUSSELL | AKIS | 7 |
| TONNETTE | JONES | 7 |
| JONATHAN | WILLIS | 6 |
| ZOLOTTIE | BOWLING | 6 |
| ALLISON | BOWE | 5 |
| BRANDON | MORTON | 5 |
| BRYAN | GASTON | 5 |
| DARRYL | DAWKINS | 5 |
| DOMINIQUE | SCALZETTI | 5 |
| ERNEST | JONES | 5 |
| HECTOR | ESCALERA | 5 |
| KEVIN | KREUSER | 5 |
| TESA | NEWTON-HART | 5 |
| YESSENIA | RIVERA | 5 |
| AARON | CAMPBELL | 4 |
| EDGAR | NIETO | 4 |
| FRANCISCO | ADAME | 4 |
| GEORGE | COX | 4 |
| JOSEPH | NAEGELE | 4 |
| JOY | KREUSER | 4 |
| KEITH | WHEELER | 4 |
| MICHAEL | MUHAMMAD | 4 |
| MICHAEL | SHOVEN | 4 |
| PAMELA | HUDSON | 4 |
| RANCE | HOPKINS | 4 |
| ROSALINDA | BANUELOS | 4 |
| TANDRA | TYLER | 4 |
| VICTOR | JUNIOUS | 4 |
| ANTWAN | JONES | 3 |
| CEDRIC | BELL | 3 |
| GEORGANNE | STRUSS | 3 |
| JAMES | SMITH | 3 |
| JASON | SMITH | 3 |
| JOHANNA | ALMARAZ | 3 |
| JUAN | ARGUELLES | 3 |
| LATERRIAN | HILL | 3 |
| MICHELLE | BAILEY | 3 |
| NICHOLETTE | VARGAS | 3 |
| RODNEY | PURDY-BLAKE | 3 |
| ROLANDO | GALINDO | 3 |
| SHANNON | CARROLL | 3 |
| THEODIS | CHAPMAN | 3 |
| TYRONE | HUTSON | 3 |
| VICKI | JONES | 3 |
| YVONNE | PULIDO | 3 |
| AMY | O'ROURKE | 2 |
| ANTHONY | LEWIS | 2 |
| BONNI | BOHL | 2 |
| CRYSTAL | DOMINGUEZ | 2 |
| DEDRICK | ROBERTS | 2 |
| DIEGO | RODRIGUEZ | 2 |
| GABRIEL | REDIC | 2 |
| JULIE | PERFETTI | 2 |

Or we can get a graph of clients by Zip code, diveded by gender.

## Clients by Gender Bar Plot



There are some issues to work out with this approach to the department stats. This include":"

- Formatting the report so it meets department standards.
  - This will take time, but formatting these reports is yet-another-language to learn
- Cleaning the data
  - Access to JEMS and the JEMS Backend is essential to do this work
  - Data will need to be cleaned, per the previous example
  - The department should work to establish better data storage practices
- Free software, the R package and RStudio, need to be installed on machines
  - Learning R is not difficult, but, it is necessary.
  - Other free software packages may need to be installed for PDF creation.

Despite these issues, I believe that statistics and data can still collected, cleaned, and transformed into useful information, regardless of stafffing concerns. Shifting away from Excell based analysis also opens up additional opportunties for data analysis, including text-mining, which can help in QA, without any increase to capital or operating expenses.