LibStats: There and Back Again

Presented By: Elizabeth Peele Mumpower, Systems Librarian
Alex Cooper, Lead Applications Analyst
Lisa Hamlett, Applications Analyst II

EMORY
LIBRARIES &
INFORMATION
TECHNOLOGY

Road Map

- **Introduction**
- **Part I:**
  - **Why this project?**
  - **How did we start it?**
- **Part II:**
  - **Engaging With the ezPaarse Open Source Community**
  - **Making ezPaarse Your Own**
- **Part III:**
  - **Using Tableau to compile and present your unique data**
  - **Some examples of our data**

Part 1. Beginning the journey

Origin, project management, and architecture

The Origins of LibStats

Have

- COUNTER
- Unparsed proxied logs
- Proxy both on & off campus

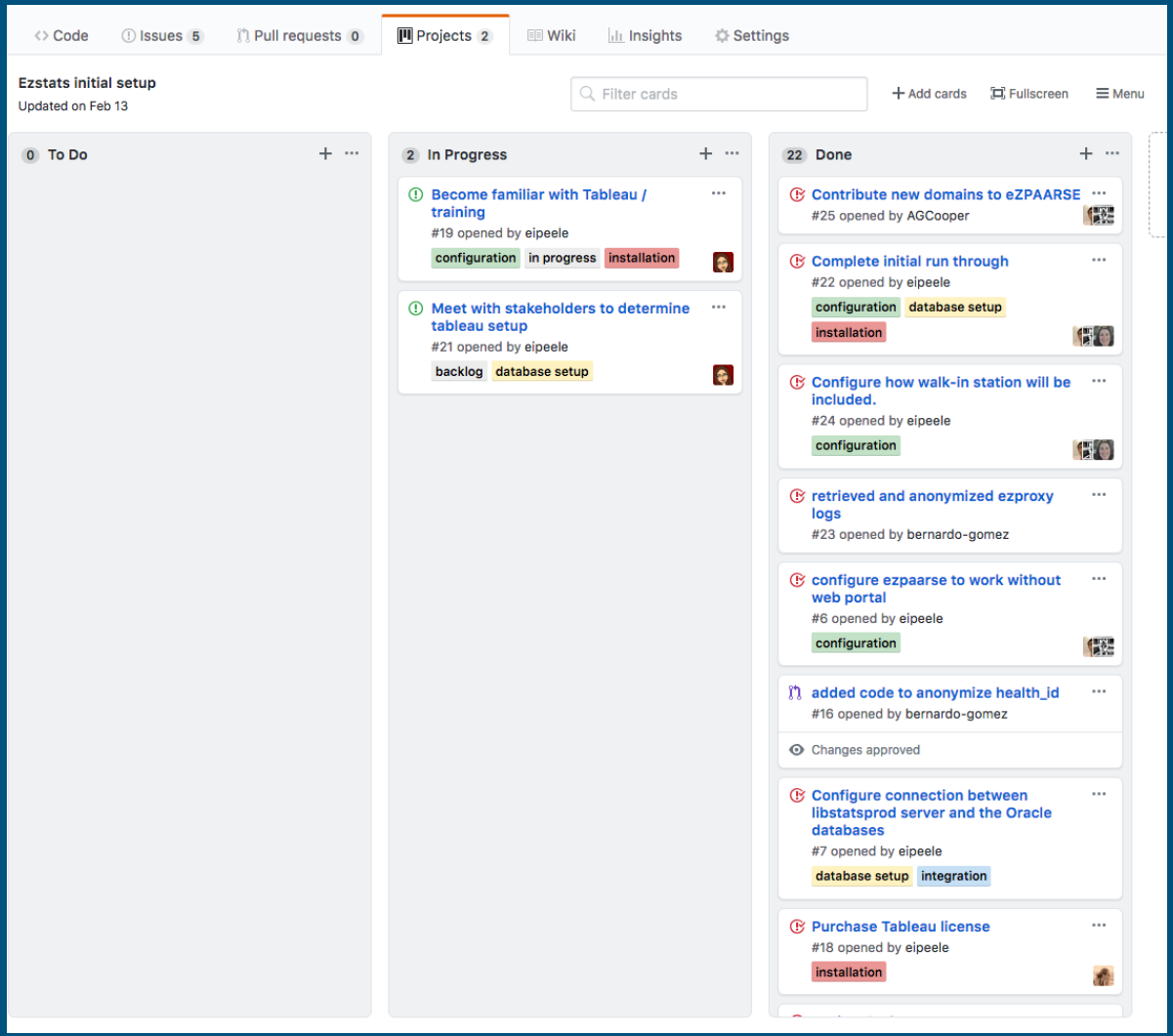
Need

- Data from resources without COUNTER
- Easier way to visualize stats
- Way to match access with demographic data for a more well-rounded view

ezPAARSE

Project Management

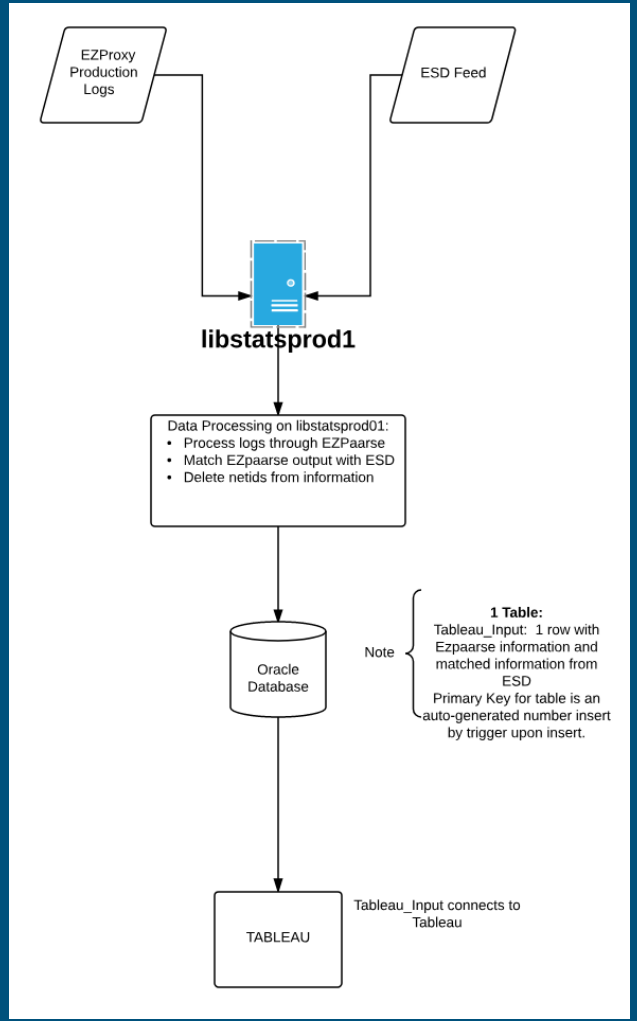
- Multiple moving components
  - Setting up ezPaarse
  - Architecting data loads
  - Designing & Implementing database
  - Setting up Tableau
  - Ensuring communication with stakeholders
  - Prioritization
  - etc
- Use GitHub to track everything
  - Private repo for initial setup and changes/tweaks to architecture
  - Public repo forked from ezpaarse-platforms



Private Repo

Architecture

- Extract:
  - Different data sources
- Transform:
  - Filter
  - Match
  - Anonymization
  - Date transformation
- Load:
  - Oracle Database



Part 2.
On the Road

Engaging with the ezPaarse community and making ezPaarse your own

Tracking the State of ezPaarse

- Couperin and the project members
- Trello: ezPaarse / AnalogIST
  - https://trello.com/b/wEaLnz8d/ezpaarse-analogist-suivi-des-plateformes
- The Couperin project AnalogIST webpage
  - http://ang.couperin.org/

Contributing to ezPaarse Part 1

- Analyze your EZProxy logs
- Create a Trello card for the platform
- Add links from the EZProxy logs to AnalogIST

Contributing to ezPaarse Part 2

- Creating your own parsers
- Use platform-init to create your platform’s parser
- Modify the parser.js template to suit your needs
- Build out your test.csv
- Commit your code to ezPaarse on GitHub
  - https://github.com/ezpaarse-project/ezpaarse-platforms
- Since 01/2018 we have added 53 parsers locally and 35 globally
  - NOTE: We do not reparse old logs when we add a new parser

Tools

- regex
  - Tooltip: https://regex101.com/
- JavaScript
- ezPaarse community

Part 3.
Telling Your Story

Using Tableau to compile and present your unique data

Preparing the Data Connection

- Tableau Public vs. Tableau Desktop
  - Same functionality, but access to data in Tableau Public cannot be restricted and Tableau owns the data
- Connect Tableau Desktop to Tableau Server
- Connect Tableau Desktop to Database with Tableau Connector
- Publish data source to Tableau Server
- Schedule a regular extract (not a live connection) of the data source
- Create a workbook connected to data source
  - Pro tip: You can connect multiple workbooks to a single data source from the server
- Publish workbook to Tableau Server

LibStats vs. COUNTER

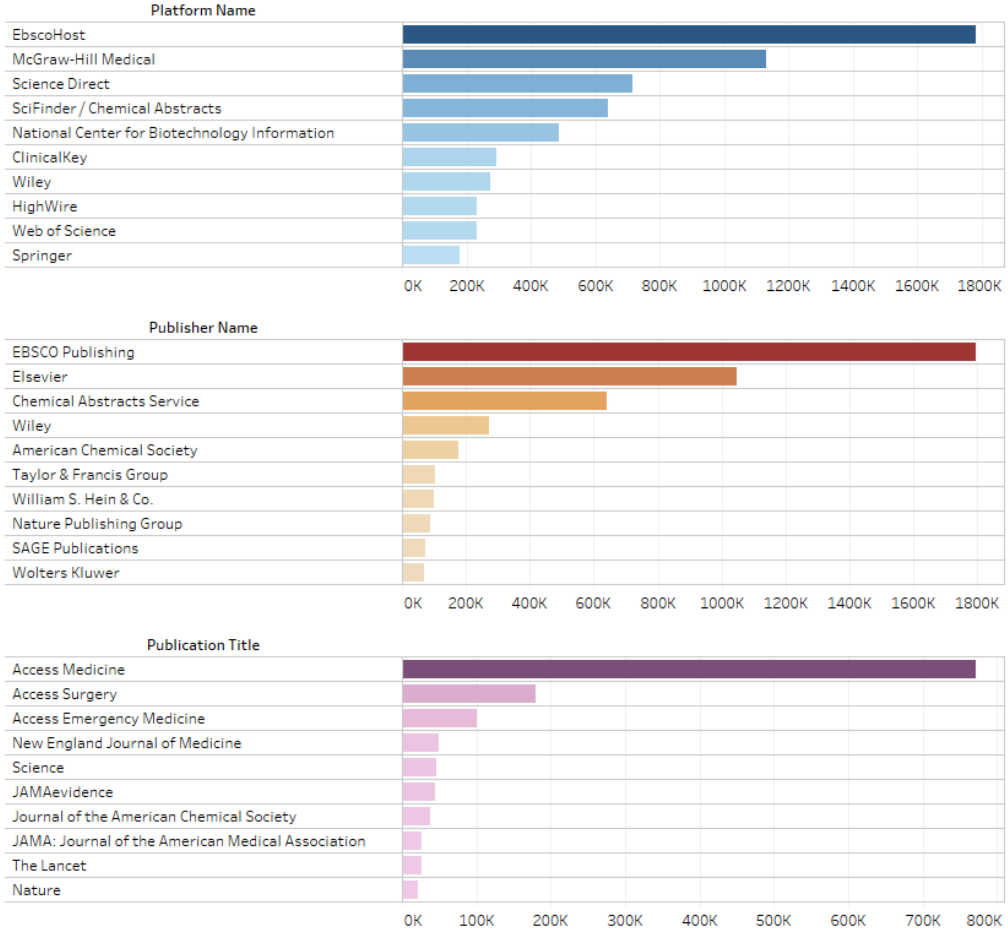
- LibStats is *not* a replacement for COUNTER reports
  - COUNTER data describe *what* resources are being accessed
  - LibStats data describe *who* is accessing the resources
- LibStats data for resources is inconsistent
  - Every website handles links and resources uniquely
  - Parser writers have different ideas of what constitutes a certain type of resource
- Both LibStats and COUNTER reports are needed to get a full picture of collection usage

LibStats Data

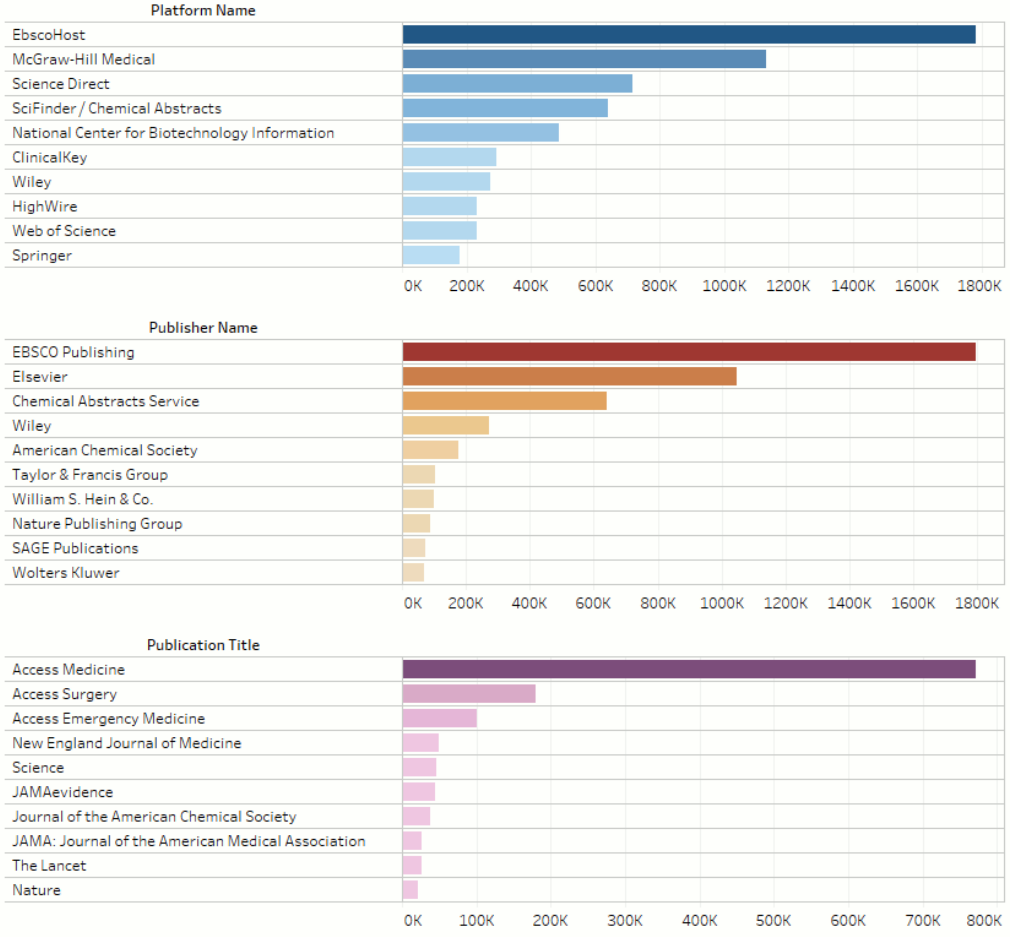
- LibStats looks at these specific data:
  - Actions, Data (Bytes), Unique Sessions, Unique Title IDs, Unique Unit IDs, & Unique User IPs
- Some data is incomplete because many parsers only show initial access
- The most useful data defined
  - Actions: Anything a user does (click, open a file, view a TOC, etc.)
  - Unique Sessions: Number of unique logins into ezProxy
  - Unique User IPs: Number of unique IP addresses that have been logged into ezProxy

That’s nice and all…
But let’s see the data!

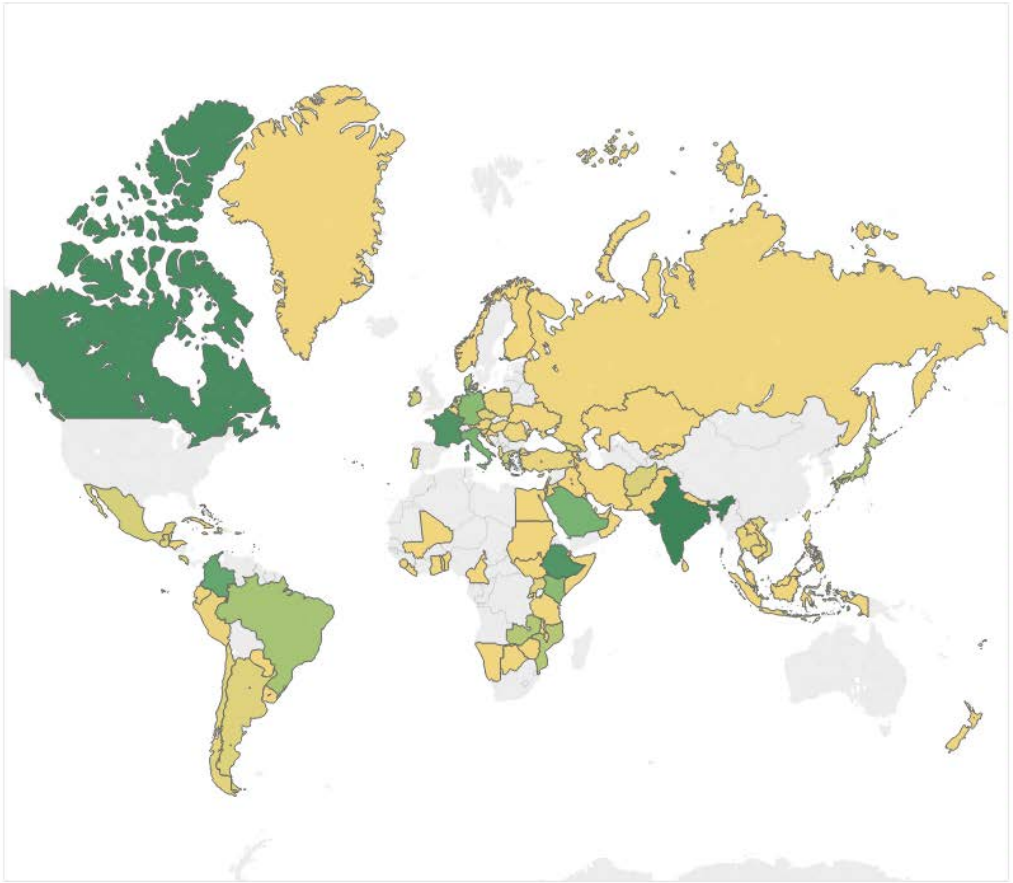
Top 10 Platforms, Publishers, and Publication Titles by Actions



Top 10 Platforms, Publishers, and Publication Titles by Actions



Click on a country either on the map or in the list to filter by that country. Click again to remove the filter.

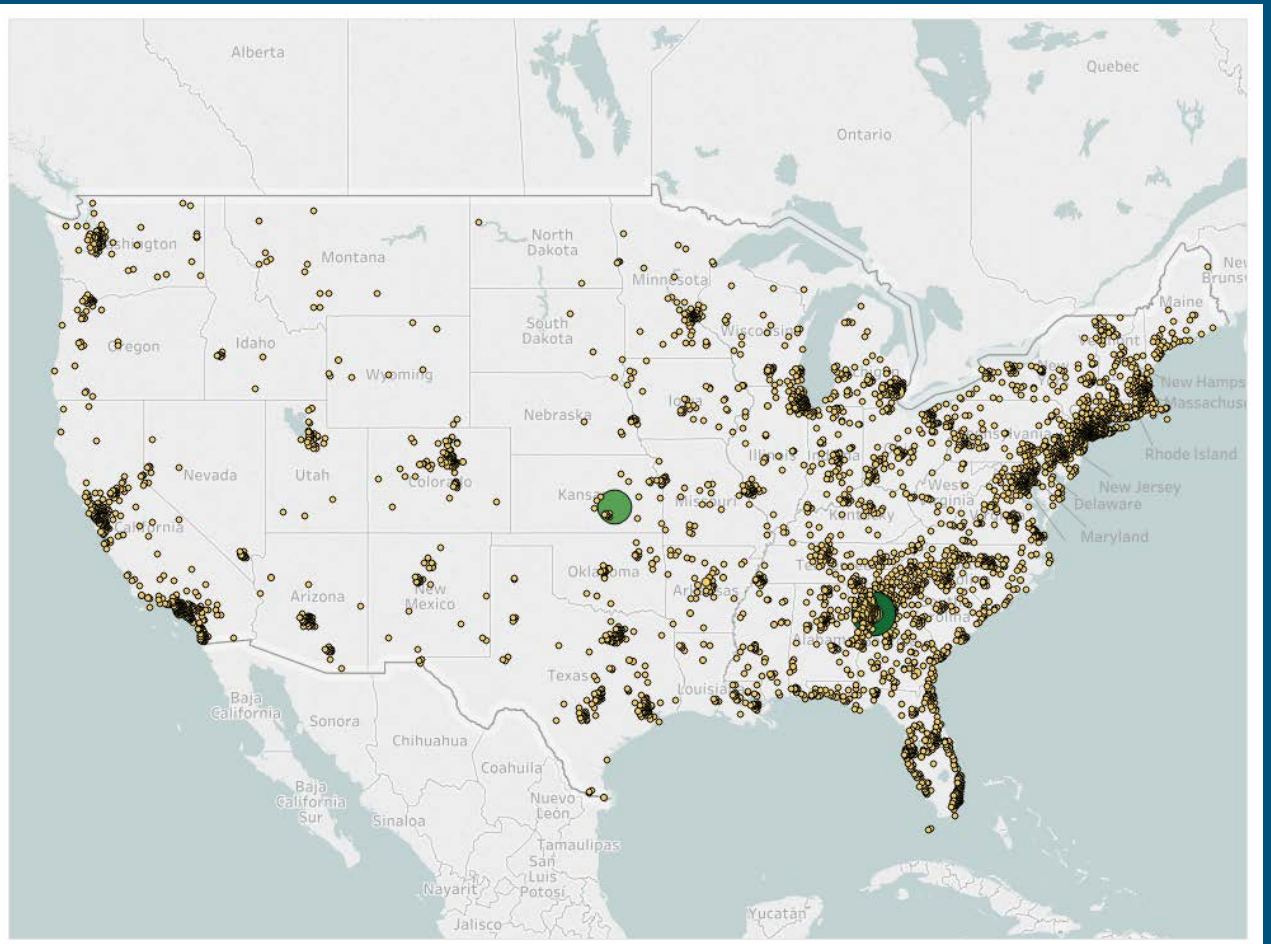


AE	349
AF	1,019
AL	21
AM	19
AR	816
AT	422
AW	2
BB	36
BE	210
BJ	50
BQ	6
BR	3,159
BS	100
BW	72
BZ	51
CA	9,461
CL	653
CM	67
CO	6,656
CR	907
CU	21
CY	4
CZ	898
DE	4,426
DJ	4
DK	2,439
EC	150
EG	72
ET	8,547
FI	21
FJ	14
FR	8,626
GD	1
GE	992
GH	117
GL	31
GR	1,627

Select Measure
Actions

Country
Outside US

Recorded Date
10/1/2016 12:00:00 AM to 6/30/2018 12:00:00 AM and Null values

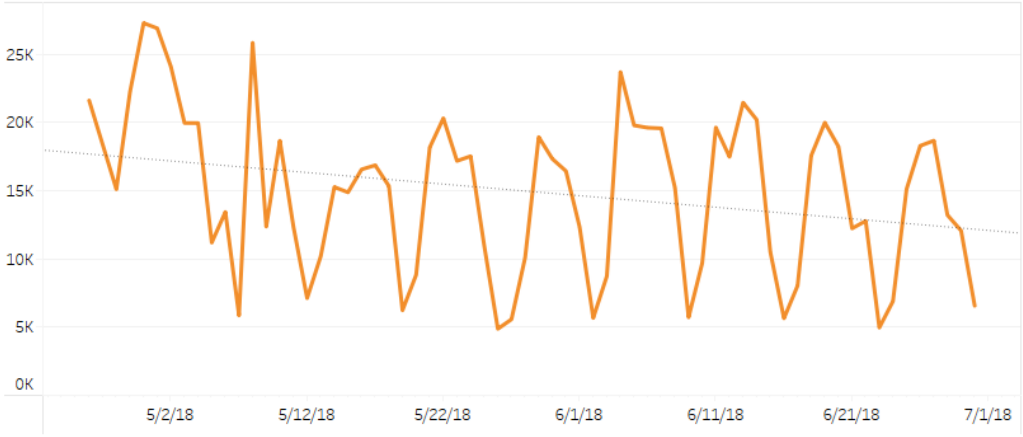


Select Measure
Actions

Country
US

Recorded Date
10/1/2016 12:00:00 AM to 6/30/2018 12:00:00 AM
and Null values

Actions by Day

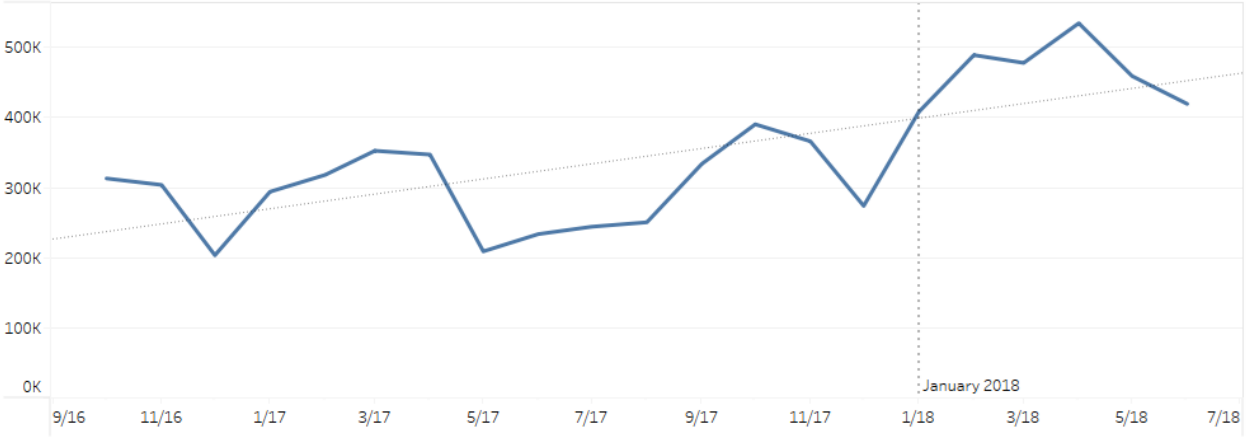


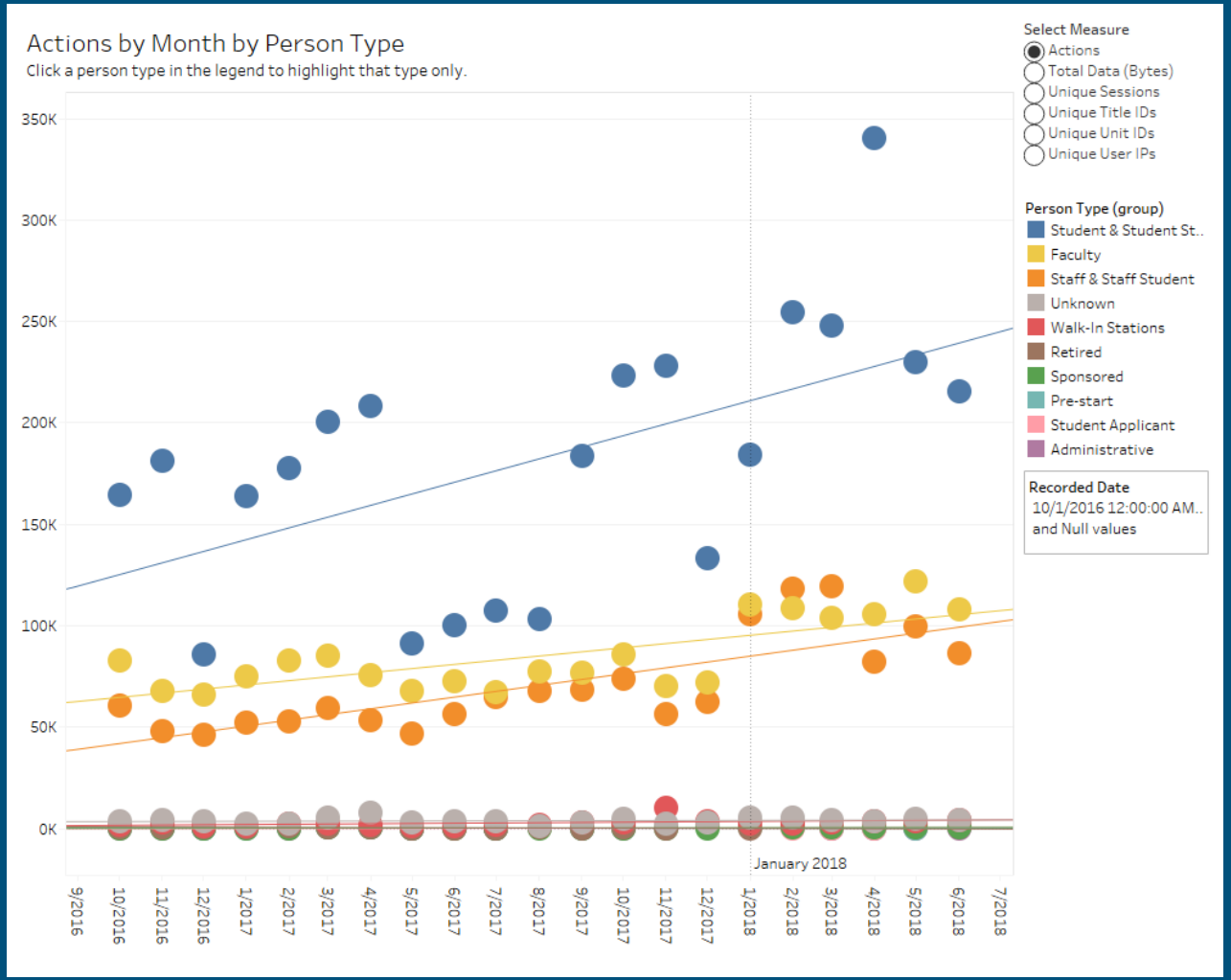
Select Measure
- (●) Actions
- ( ) Total Data (Bytes)
- ( ) Unique Sessions
- ( ) Unique Title IDs
- ( ) Unique Unit IDs
- ( ) Unique User IPs

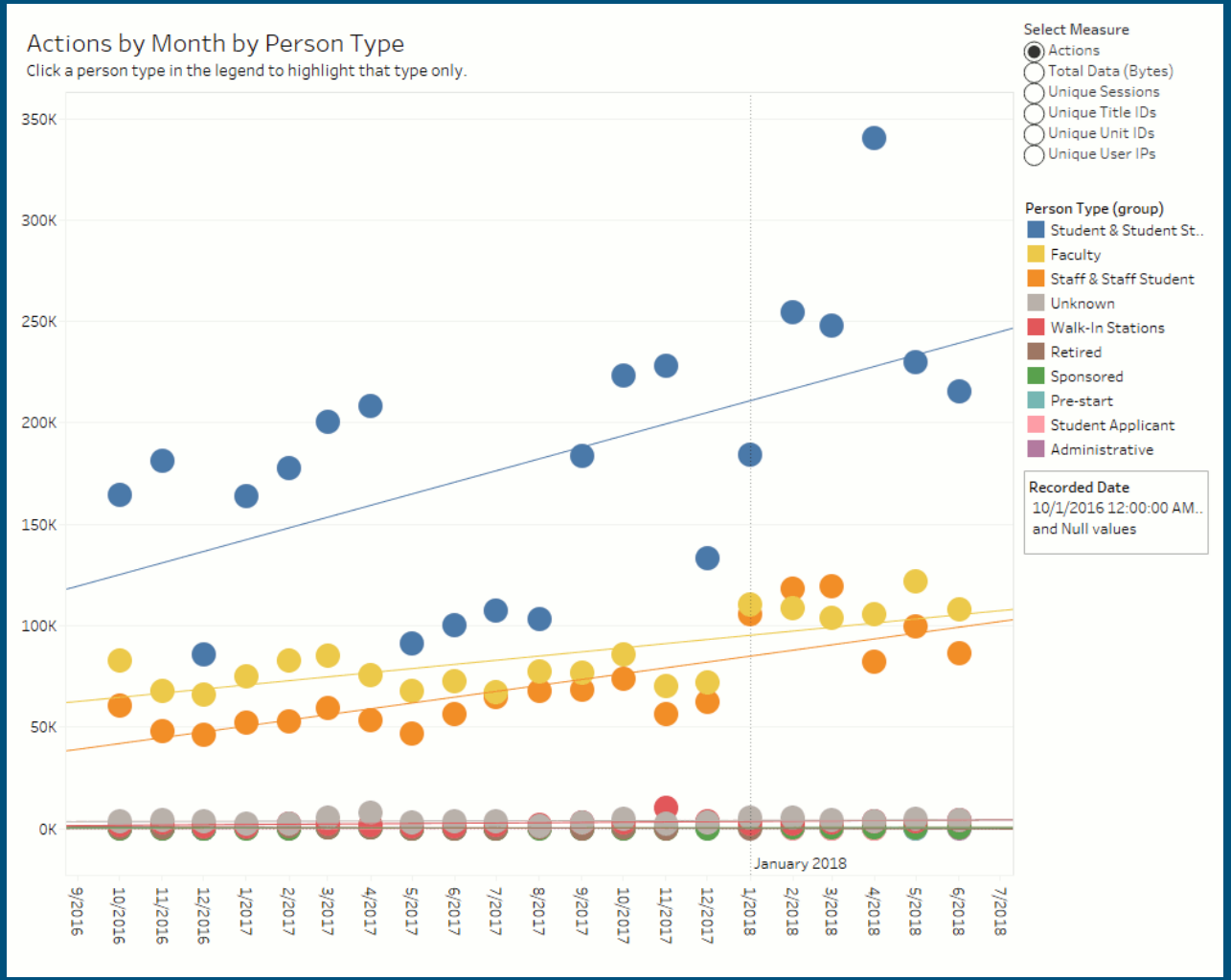
Range of Dates
April 26, 2018 to July 24, ..

Recorded Date
10/1/2016 12:00:00 AM to..
and Null values

Actions by Month

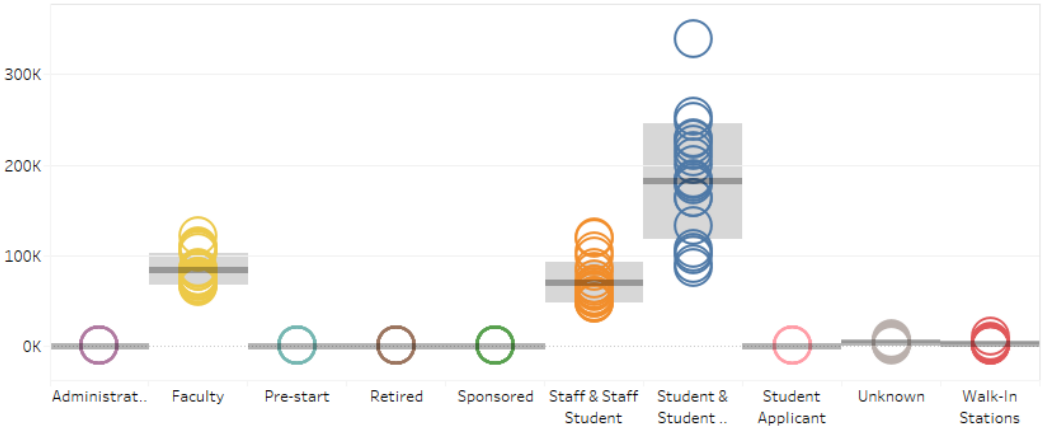



Actions by Month by Person Type
Click a person type in the legend to highlight that type only.
350K
300K
250K
200K
150K
100K
50K
0K
January 2018
9/2016
10/2016
11/2016
12/2016
1/2017
2/2017
3/2017
4/2017
5/2017
6/2017
7/2017
8/2017
9/2017
10/2017
11/2017
12/2017
1/2018
2/2018
3/2018
4/2018
5/2018
6/2018
7/2018
Select Measure
Actions
Total Data (Bytes)
Unique Sessions
Unique Title IDs
Unique Unit IDs
Unique User IPs
Person Type (group)
Student & Student St..
Faculty
Staff & Staff Student
Unknown
Walk-In Stations
Retired
Sponsored
Pre-start
Student Applicant
Administrative
Recorded Date
10/1/2016 12:00:00 AM..
and Null values

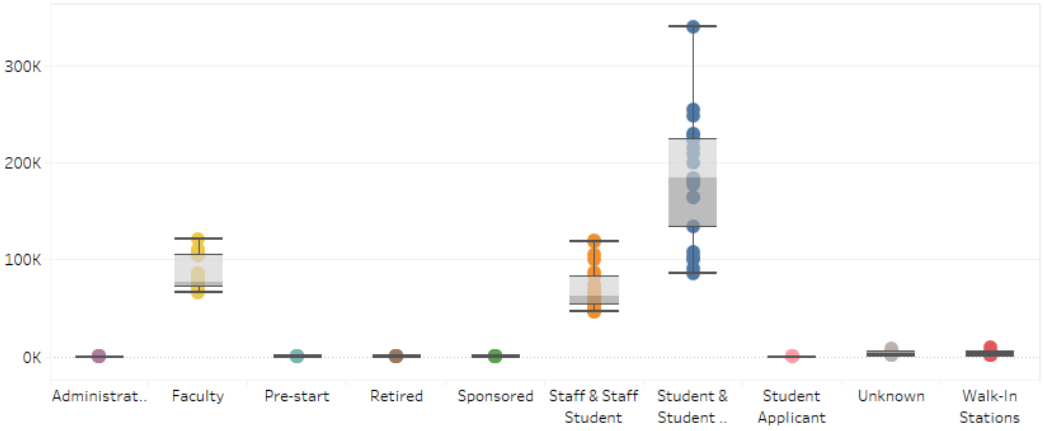

Actions by Month by Person Type
Click a person type in the legend to highlight that type only.
350K
300K
250K
200K
150K
100K
50K
0K
January 2018
9/2016
10/2016
11/2016
12/2016
1/2017
2/2017
3/2017
4/2017
5/2017
6/2017
7/2017
8/2017
9/2017
10/2017
11/2017
12/2017
1/2018
2/2018
3/2018
4/2018
5/2018
6/2018
7/2018
Select Measure
Actions
Total Data (Bytes)
Unique Sessions
Unique Title IDs
Unique Unit IDs
Unique User IPs
Person Type (group)
Student & Student St..
Faculty
Staff & Staff Student
Unknown
Walk-In Stations
Retired
Sponsored
Pre-start
Student Applicant
Administrative
Recorded Date
10/1/2016 12:00:00 AM..
and Null values

Actions by Person Type

Solid lines are averages and grey bands are standard deviation away from average



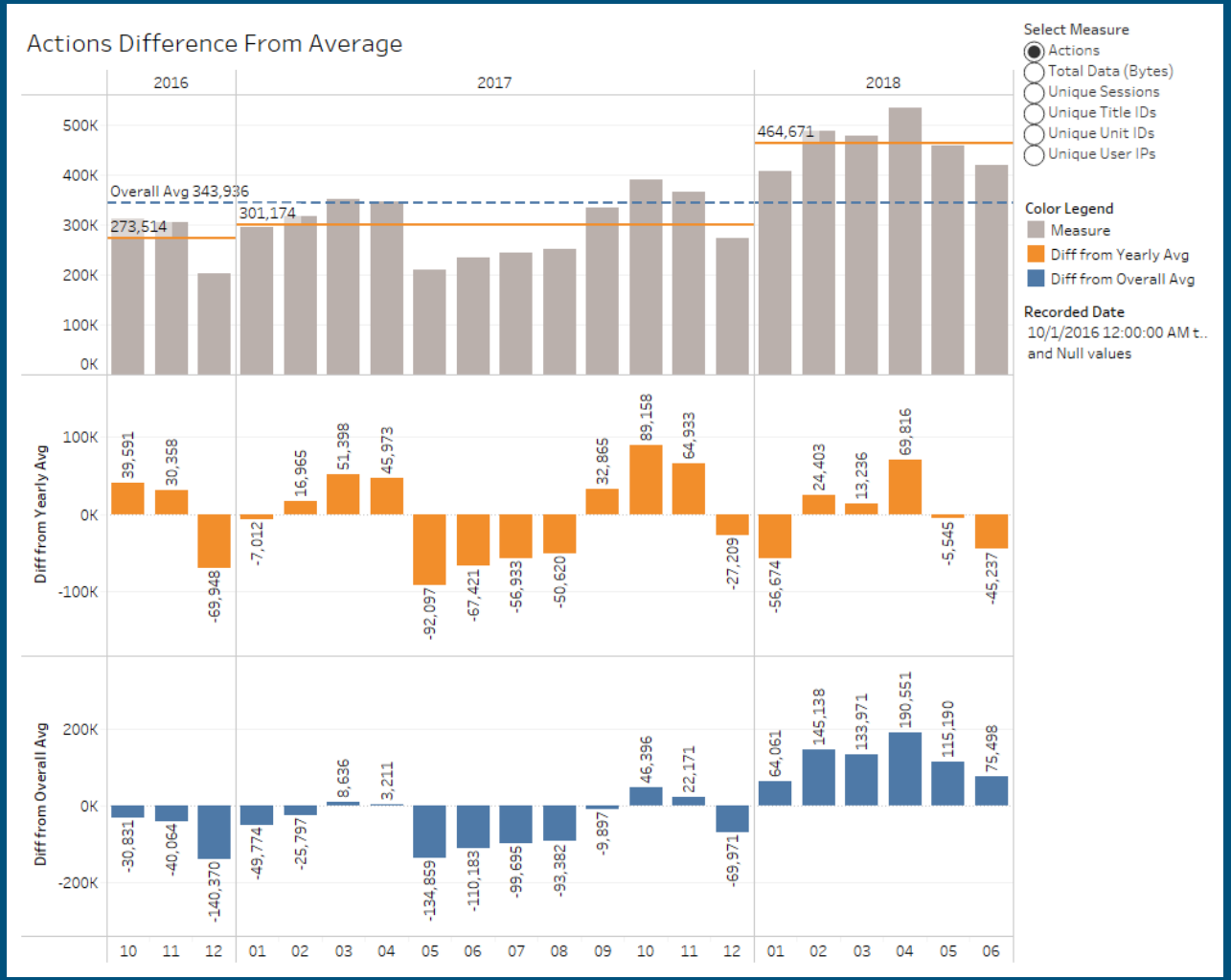
Hover over box plot to see more details



Select Measure
- Actions
- Total Data (Bytes)
- Unique Sessions
- Unique Title IDs
- Unique Unit IDs
- Unique User IPs

Person Type (group)
- Administrative
- Faculty
- Pre-start
- Retired
- Sponsored
- Staff & Staff Student
- Student & Student St..
- Student Applicant
- Unknown
- Walk-In Stations

Recorded Date
10/1/2016 12:00:00 AM.. and Null values


Actions Difference From Average
2016
2017
2018
Overall Avg 343,936
273,514
301,174
464,671
500K
400K
300K
200K
100K
0K
Diff from Yearly Avg
100K
0K
-100K
39,591
30,358
-69,948
-7,012
16,965
51,398
45,973
-92,097
-67,421
-56,933
-50,620
32,865
89,158
64,933
-27,209
-56,674
24,403
13,236
69,816
-5,545
-45,237
Diff from Overall Avg
200K
0K
-200K
-30,831
-40,064
-140,370
-49,774
-25,797
8,636
3,211
-134,859
-110,183
-99,695
-93,382
-9,897
46,396
22,171
-69,971
64,061
145,138
133,971
190,551
115,190
75,498
10
11
12
01
02
03
04
05
06
07
08
09
10
11
12
01
02
03
04
05
06
Select Measure
Actions
Total Data (Bytes)
Unique Sessions
Unique Title IDs
Unique Unit IDs
Unique User IPs
Color Legend
Measure
Diff from Yearly Avg
Diff from Overall Avg
Recorded Date
10/1/2016 12:00:00 AM t..
and Null values

Questions?

THANK YOU!

CONTACT:

Elizabeth Peele Mumpower, Systems Librarian

epeele@emory.edu

Alex Cooper, Lead Applications Analyst

acoope5@emory.edu

Lisa Hamlett, Applications Analyst II

lhamlet@emory.edu

Credits & Links

- Presentation template by SlidesCarnival
- https://trello.com/b/wEaLnz8d/ezpaarse-analogist-suivi-des-plateformes
- http://ang.couperin.org/
- https://github.com/ezpaarse-project/ezpaarse-platforms
- https://regex101.com/
- https://github.com/emory-libraries/ezpaarse-platforms