Intermediate Report, Initial Hypotheses

Information Visualization CS5764

Team Maximum Occupancy

Christopher Wakeley (chrisiw)

Patricio Moreno (pxmore91)

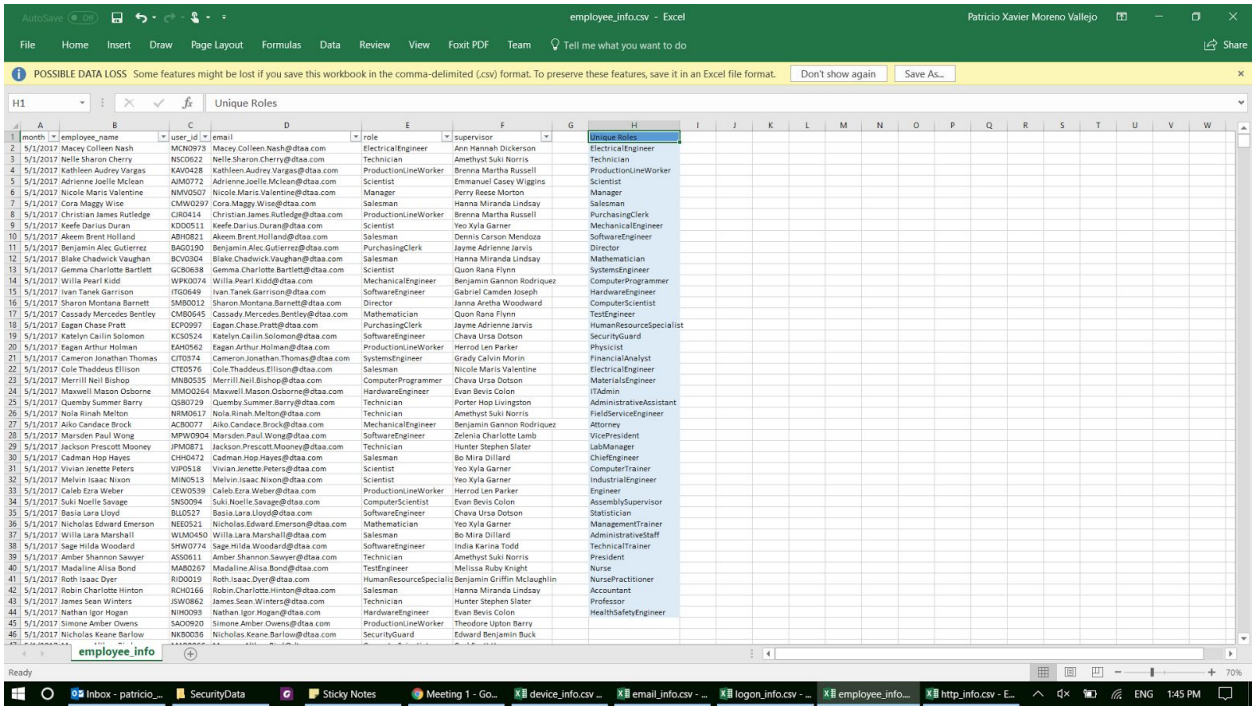
Sourabh Shetty (sour)

1. Analysis Methods

The following analysis methods are listed in chronological order.

1.1 Initial Exploration

The first step we took towards analyzing the data consisted of opening the files in a text editor to get an initial sense of the type of content we needed to process. Our next intuition was to open the files in Google sheets, but found that Google imposes a size limit which some files exceeded. Next, we tried opening the files in Excel, which was able to handle their size. In Excel, we examined the “role” attribute of the epmloyee_info file and found there are around 4-5 levels in the employee hierarchy and that it stops at the role “president”. We also noticed that the file was structured such that the full list of employees was listed for each month, leading to multiple instances of employees employed for more than a month. Lastly, we examined device_info in excel and confirmed that the “activity” attribute only consisted of the values “connect” and “disconnect”.



Unique Roles	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11
Professor	1	0	0	0	0	0	0
ManagementTrainer	1	1	1	1	1	1	1
President	1	1	1	1	1	1	1
Nurse	1	1	1	1	1	1	1
Attorney	2	2	2	2	2	2	2
AdministrativeStaff	2	2	2	2	2	2	2
ComputerTrainer	3	3	3	3	3	3	3
TechnicalTrainer	3	3	3	3	3	3	3
NursePractitioner	3	3	3	3	3	3	3
HealthSafetyEngineer	3	3	3	3	3	3	3
FinancialAnalyst	4	4	4	4	4	4	4
FieldServiceEngineer	5	5	5	5	5	5	5
HardwareEngineer	6	5	5	5	5	5	5
VicePresident	6	6	6	6	6	6	6
LabManager	6	6	6	6	6	6	6
Engineer	6	6	6	6	6	6	6
AssemblySupervisor	6	6	6	6	6	6	6
ChiefEngineer	7	7	7	7	7	7	7
IndustrialEngineer	7	7	7	7	7	6	6
Accountant	7	7	7	7	7	7	7
Statistician	8	8	8	8	8	8	8
MaterialsEngineer	9	9	9	9	9	9	9
ITAdmin	9	9	8	8	8	8	8
Director	12	12	12	12	12	12	12
HumanResourceSpeci	12	12	12	12	12	12	12
PurchasingClerk	13	13	13	13	13	13	13
SecurityGuard	15	15	15	15	15	15	15
TestEngineer	16	16	16	16	16	16	16
SystemsEngineer	17	16	16	16	16	16	16
Manager	23	23	23	23	23	23	23
ComputerProgrammer	25	26	25	25	25	26	25
ComputerScientist	28	28	28	28	28	28	28
MechanicalEngineer	29	29	29	29	29	29	28
Mathematician	30	30	29	28	28	28	28
Physicist	30	30	30	30	29	29	28
Scientist	32	32	32	32	32	32	32
AdministrativeAssistar	36	36	36	36	36	36	35
ElectricalEngineer	41	41	41	41	41	40	39
SoftwareEngineer	42	41	40	40	40	40	40
Salesman	147	146	146	145	144	141	140
Technician	148	148	147	146	146	146	145
ProductionLineWorker	178	176	173	172	171	169	167
42	980	974	966	962	959	953	944

1.2 File summaries

To get a better grasp on the data we were given, we thought it would be useful to produce tables with some summary statistics of each file. We obtained the ranges and number of unique values using text editors and Tableau.

Device_info.csv

Attribute	Unique values	Range

Date	153,896	5/1/2017 1:58:03 - 10/31/2017 6:51:21
User	215	~
PC	212	~
Activity	2	Connect(79,029), Disconnect(76,984)

employee_info.csv

Attribute	Unique values	Range	Example Value
Month	7	5/1/2017 - 11/1/2017	5/1/2017
Name	980	~	Macey Colleen Nash
User ID	980	~	MCN0973
Email	980	~	Macey.Colleen.Nash@dtaa.com
Role	42	~	ElectricalEngineer
Supervisor	68	(Includes empty value for the supervisor of the President)	Ann Hannah Dickerson

logon_info.csv

Attribute	Unique values	Range
Date	96,131	5/1/2017 1:23:00 AM - 10/31/2017 7:54:00 PM
User	979	
PC	934	
Activity	2	

email_info.csv

Attribute	Unique values	Range
Date	357,560	05/01/2017 06:46:51 10/31/2017 19:48:19

To	110,101	
From	980	
size		6,987 - 85,904
Attachments		0 - 9

http_info.csv

Attribute	Unique values	Range
Date	733,286	2017-05-01 00:00:36 2017-10-31 19:17:58
User	978	
PC	931	
url	4116	

1.3 Employee Hierarchy Tool

We created a tool that took in the employee_info.csv file and parsed it into a more navigable employee hierarchy tree structure.

This tool lists the employees relative to their position in the organization for each month, finds an employee's immediate subordinates and all their superiors, and lists all the employees that worked at the organization for under seven months.

```
1. President: Kaden Tarik Giles
  2. VicePresident: Janna Aretha Woodward
    3. Director: Sharon Montana Barnett
      4. ComputerTrainer: Oleg Ahmed Porter
      4. ManagementTrainer: Ezra Chadwick Sargent
      4. TechnicalTrainer: Meredith Carly Richardson
      4. ComputerTrainer: George Rahim Cross
      4. ComputerTrainer: Arden Chaim Cohen
      4. AdministrativeAssistant: Gil Walker Gaines
      4. TechnicalTrainer: Rebecca Rhiannon Battle
      4. TechnicalTrainer: Georgia Hope Shaffer
    3. Director: Rudyard Talon Hickman
      4. Manager: Edward Benjamin Buck
        5. SecurityGuard: Nicholas Keane Barlow
        5. SecurityGuard: Cassandra Francesca Buchanan
        5. SecurityGuard: Danielle Ila Benson
        5. SecurityGuard: Ahmed Merrill Rodgers
        5. SecurityGuard: Kai Alexandra Carroll
        5. SecurityGuard: Thomas Luke Gonzales
        5. SecurityGuard: Amela Hannah Sweeney
        5. SecurityGuard: Jeremy Logan Hill
        5. SecurityGuard: Mikayla Sacha Figueroa
        5. SecurityGuard: Idona Amela Frederick
        5. SecurityGuard: Amela Libby Bird
        5. SecurityGuard: Hall Curran Hood
        5. SecurityGuard: Ria Hiroko Barnes
        5. SecurityGuard: Hadassah Alma Prince
        5. SecurityGuard: Wylie Keane Conway
      4. AdministrativeAssistant: Nicholas Bert Pittman
      4. Manager: Jaquelyn Yoshi Mckinney
        5. ITAdmin: Zelenia Neve Snider
        5. ITAdmin: Breanna Leila Alexander
        5. ITAdmin: Uta Adele Powers
        5. ITAdmin: Hashim Kasper Woodard
        5. ITAdmin: Lucius Ulysses Phelps
        5. ITAdmin: Whilemina Shafira Perry
        5. ITAdmin: Jerry Vincent Shepherd
        5. ITAdmin: Daquan Wing Gutierrez
        5. ITAdmin: Prescott Anthony Whitley
    3. Director: Anika Judith Valentine
      4. HealthSafetyEngineer: Akeem Paki Chapman
      4. HealthSafetyEngineer: Hedy Yvette Cook
      4. HealthSafetyEngineer: Benjamin Nash Boone
    3. Director: Ryder Ivan Bowens
```

```
Palmer Nathaniel Church, Manager
Immediate Subordinates:
        Sybil Eliana Britt
        Lucas Grady Wyatt
        Norman Levi Bentley
        Geraldine Linda Warner
        Justin Declan Gross
        Bree Isabelle Sweet
Superiors:
        Buffy Meredith Giles
        Kaden Tarik Giles
```

1.4 Logon activity by PC

Using Tableau and logon_info.csv as source, we aggregated the users by PC to visualize the PCs with more logons and logoffs. We discovered that PC-7165 has the greatest amount of logon activity than others PCs, as shown in figure 1, while the remaining PCs are almost similar in terms of logon activity.

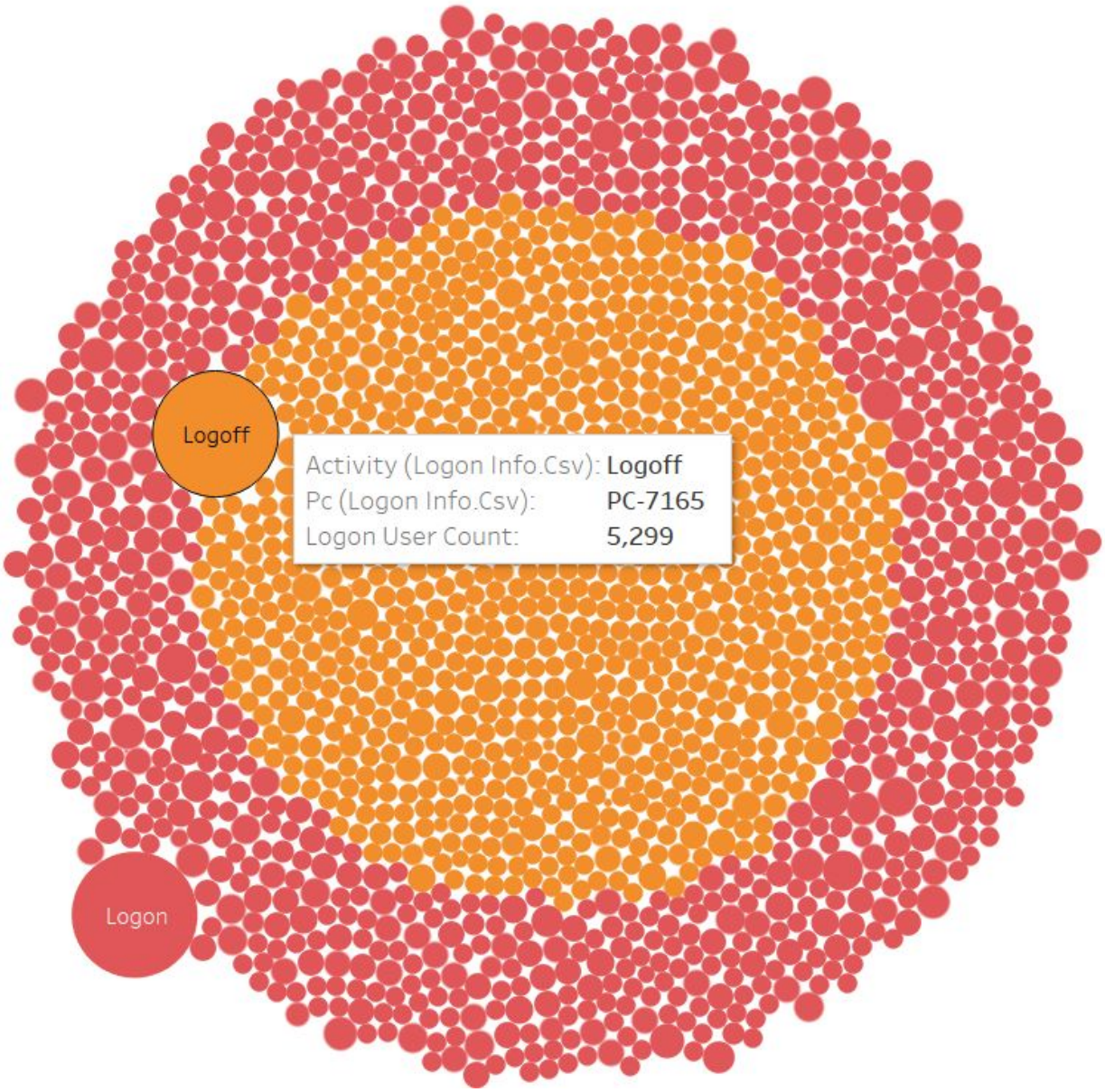


Figure 1. Logon Activity by PC

To determine if PC-7165 is used by only one users or many users, we created a tree map (figure X) where each PC is the parent of the users who have logon activity on it. The PC-7165 is the biggest rectangle in figure 2. As we can see, this PC is used by many users. So we decided that PC-7165 is a shared PC maybe in a public space that is not longer relevant for the analysis .

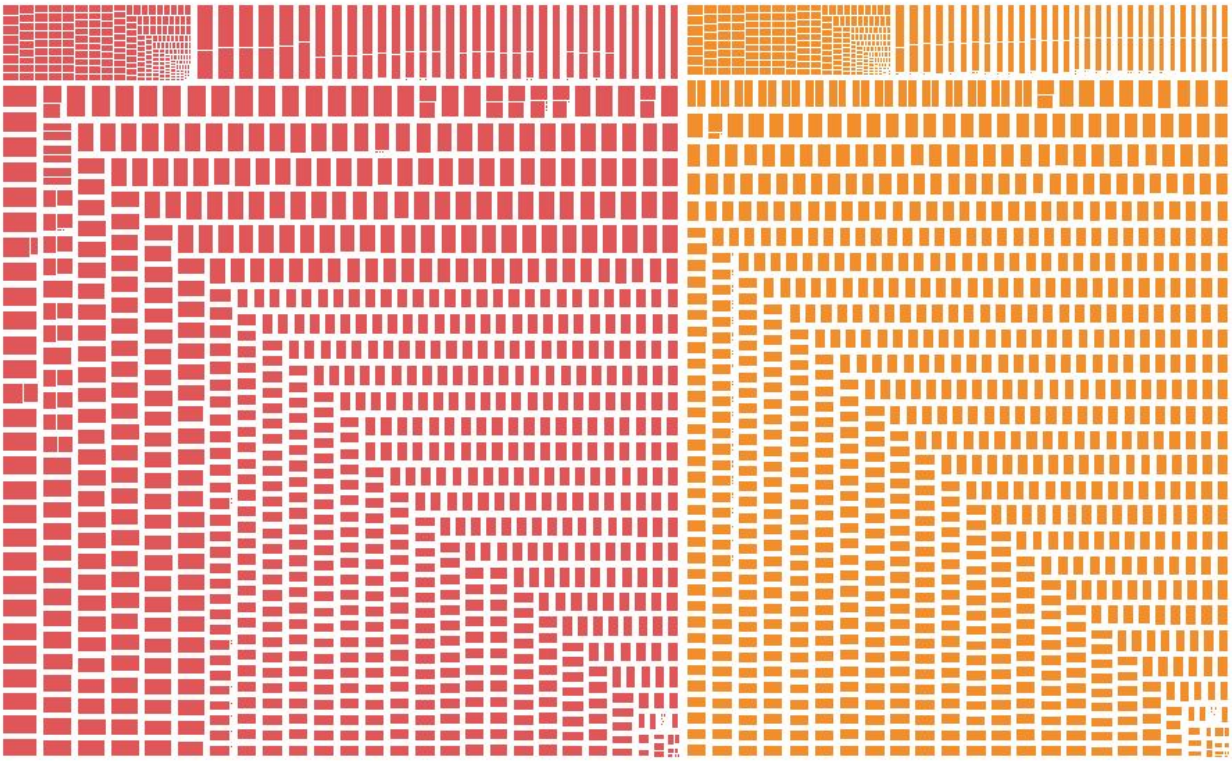
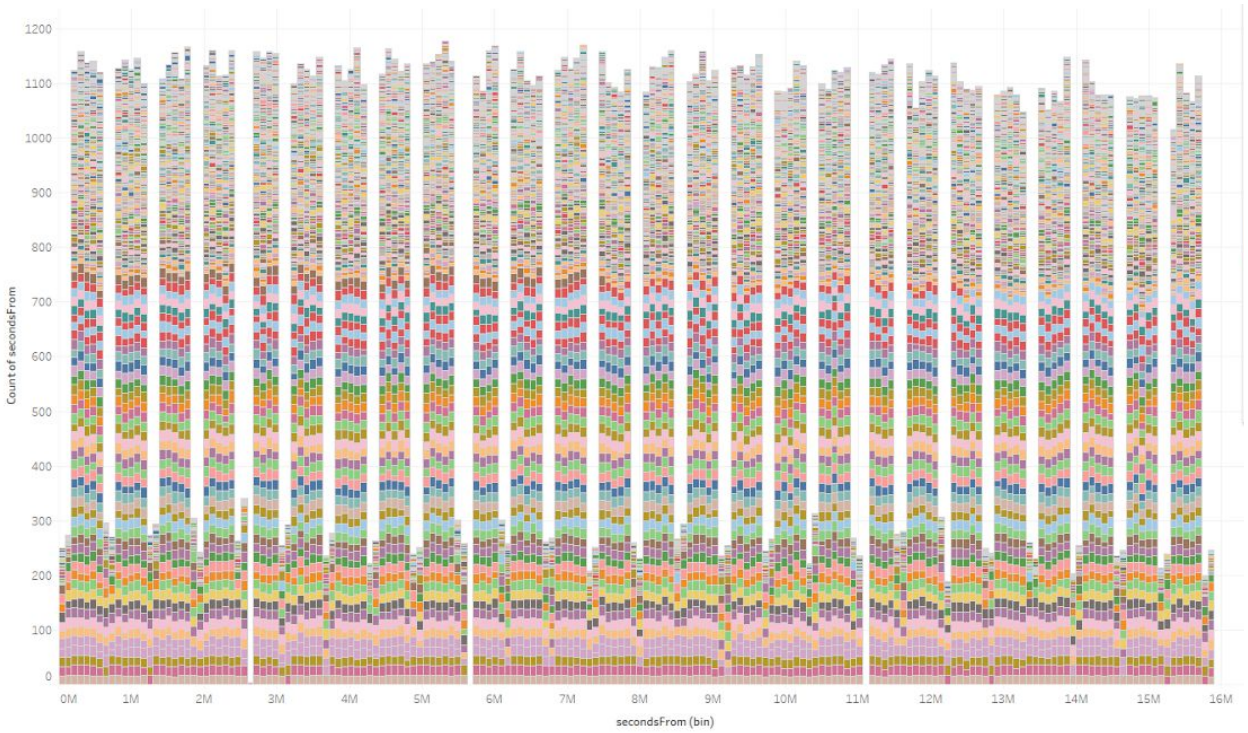


Figure 2. Tree map of PCs and user activity.

1.5 device_info.csv Gantt Chart/Histogram

After initial exploration, we had the intuition that the frequency of certain events might show some anomalous behavior such as logon events or device activity events. To this end, we created a Tableau sheet to visualize the device_info file entry events as a histogram. We created a calculated field that represents the number of seconds an event took place after the first entry in the file and binned this measure by intervals of 86400 seconds, or the number of seconds in a day after playing with the bin size a bit. Next, we tried dragging the User dimension onto the chart and Tableau automatically produced the Gantt chart below. We sorted the stacked bars by size with the largest bars on the bottom of the graph. The first thing we noticed was that there was a consistent set of employees with the highest device activity, indicated by the horizontal colored bands. We also noticed the cyclic pattern of breaks in activity, most likely attributed to weekends, and the less frequent complete breaks in activity which we initially suspected were holidays, but have since decided to reexamine. Lastly, we noticed one color band abruptly stopped around the 13M second mark. It seemed to be the only employee with high device activity to be fired. After later corroboration with other files, we added this employee to our suspect list.



1.6 Website Visitation

We extracted the domains from the urls to know how many times a domain is visited. For this, we calculated the number of times a domain appears in the dataset and then we created a packed bubble visualization(figure 3). We noticed there are 394 domains in the http dataset and the most visited are Youtube, Twitter, Yahoo, Google, Facebook, Target, etc. However, we will focus on the domains that are not commonly visited to find any suspicious activity.

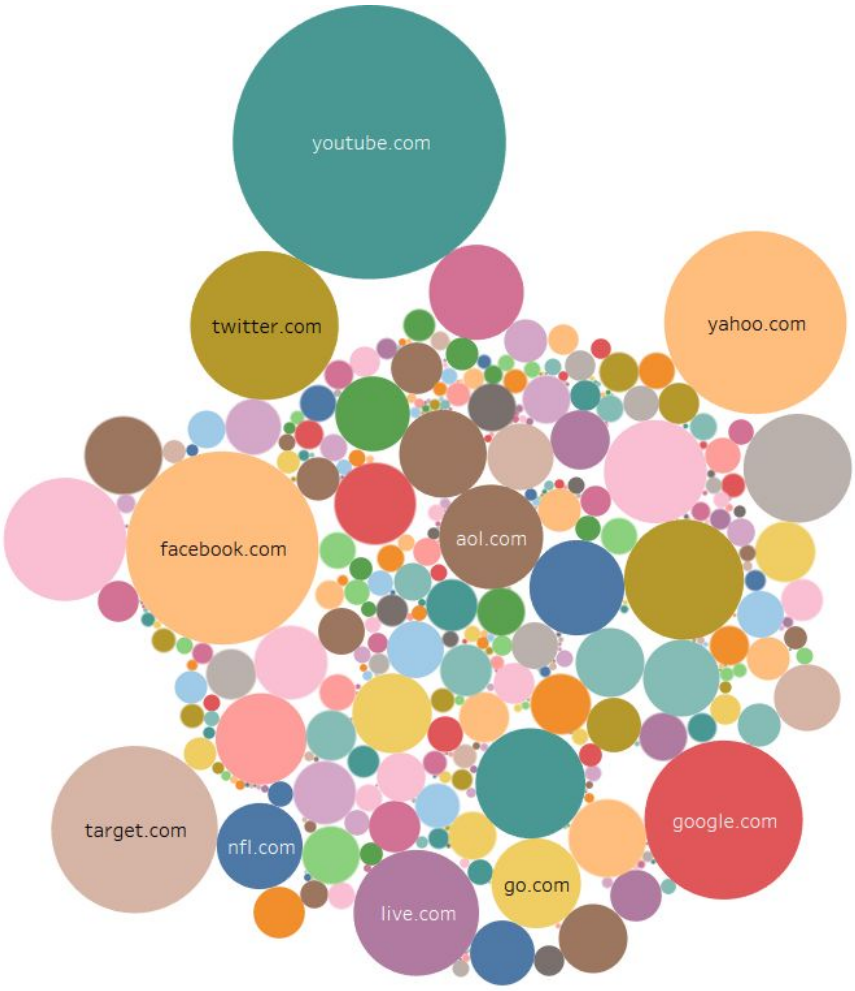


Figure 3. Visualization of number of visits per domain

1.7 Merging files with SQL

The data was loaded in a mysql database. Each table in the database corresponds to a csv file as shown in figure 4. We used sql queries to ease the location of suspicious records in thousands of entries and get ranges of data based on dates. Moreover, Mysql was useful to link data in different tables using joins. For instance, we joined the table email with employee. While joining these tables we found that there is an email sent from an external email address ("Chris.L.North@vt.edu") not linked to any employee. This finding is explained in more detail in section 2.

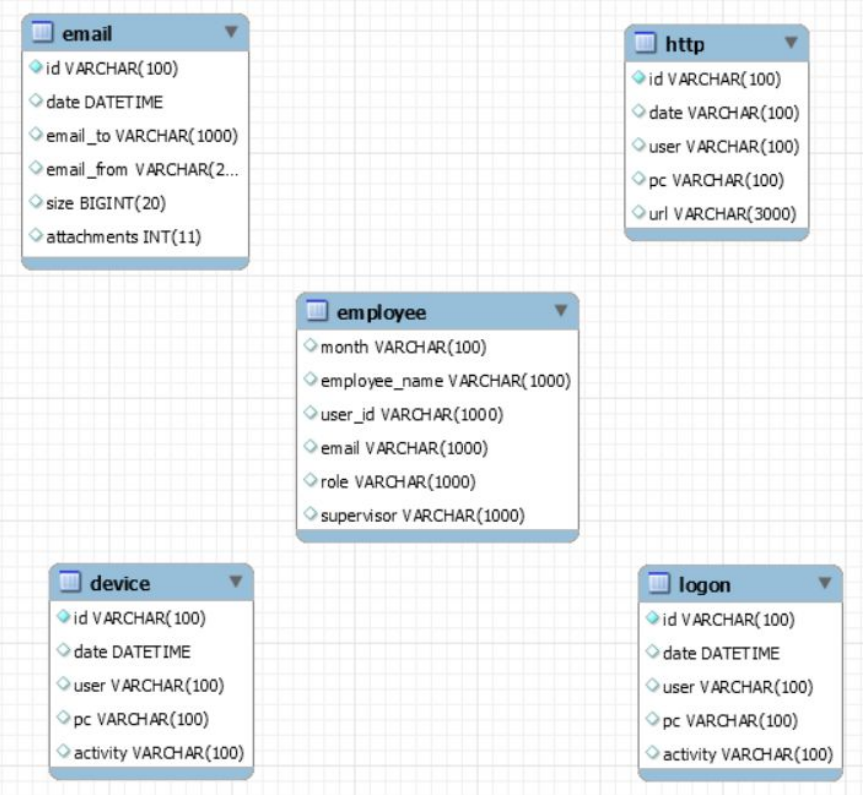
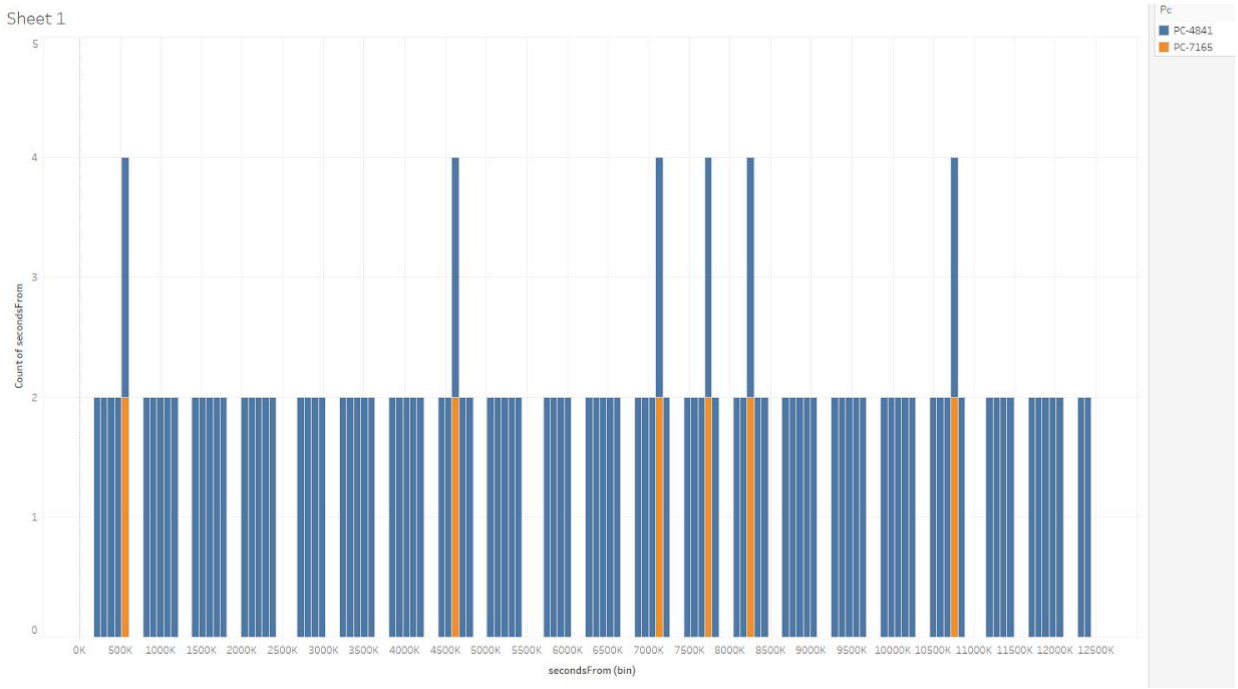


Figure 4. MySql Database

1.8 Histograms of logon_info.csv

After compiling an initial list of suspicious people, we thought of ways to analyze their behavior to find new avenues of investigation or confirm/reject them as a person of interest. One such analysis was creating a histogram from the logon activity in order to illuminate any anomalous patterns. We created these histograms much like the Gantt chart in section 1.5, however we added an additional filter for a specific employee on the User dimension. Below is an example histogram showing an employee who logged on consistently twice a day with the exception of when they logged into a shared computer a handful of times during their time at the company.

So far this analysis method has not lead to anything concrete.



1.9 Fired Employee Queries

```
Fired Employees:
Months Worked   Last Month                 Name
    6            10/01/2017            Simone Amber Owens
    6            10/01/2017            Megan Brynn Snyder
    2            06/01/2017            Scarlett Kerry Charles
    6            10/01/2017            Damon Hyatt Boyle
    1            05/01/2017            Ruth Daphne Carver
    3            07/01/2017            Palmer Ahmed Sutton
    3            07/01/2017            Chaney Sean Fuentes
    2            06/01/2017            Giacomo Hamilton Giles
    6            10/01/2017            Dane Bert Mcfadden
    6            10/01/2017            Halee Morgan Wilkins
    5            09/01/2017            Jada Quinn Vang
    4            08/01/2017            Paki Jason Emerson
    6            10/01/2017            Hector Martin Daugherty
    3            07/01/2017            Lucian Rajah Lloyd
    5            09/01/2017            Levi Keegan Foley
    1            05/01/2017            Sonia Nerea Love
    2            06/01/2017            Ima Quintessa Williams
    1            05/01/2017            Garrison George Guerrero
    2            06/01/2017            Whilemina Shafira Perry
    4            08/01/2017            Justine Amy Lara
    1            05/01/2017            Chris L North
    2            06/01/2017            Carson Caesar Olson
    6            10/01/2017            Shellie Medge Jefferson
    1            05/01/2017            Hanae Rhiannon Lindsey
    4            08/01/2017            Noah Perry Bolton
    5            09/01/2017            Chancellor Cody Douglas
    5            09/01/2017            Martin Jonathan Martin
    5            09/01/2017            Denise Serina Le
    3            07/01/2017            Lillith Chanda Delgado
    1            05/01/2017            Dominic Zeph Johns
    2            06/01/2017            Kenyon Arthur Kent
    5            09/01/2017            Cedric Cyrus Harrison
    6            10/01/2017            Francesca Kylie Russo
    1            05/01/2017            Halee Autumn Craft
    5            09/01/2017            Kelly Hashim Carr
    2            06/01/2017            Philip Isaiah Best
```

We upgraded our employee info tool to list those employees that worked for under seven months. This could be people that were fired, or quit right after stealing the data.

We also queried those employees using sql to sort the employees by role. We can see there are unique roles in the filtered data. Those roles are Professor, ITAdmin, MechanicalEngineer, and AdministrativeAssistant

months	user_id	employee_name	email	role	supervisor
6	SMJ0486	Shellie Medge Jefferson	Shellie.Medge.Jefferson@dtaa.com	Technician	William Vernon Booth
3	LRL0873	Lucian Rajah Lloyd	Lucian.Rajah.Lloyd@dtaa.com	Technician	Hunter Stephen Slater
2	PIB0855	Philip Isaiah Best	Philip.Isaiah.Best@dtaa.com	Technician	Hunter Stephen Slater
1	SNL0096	Sonia Nerea Love	Sonia.Nerea.Love@dtaa.com	SystemsEngineer	Evan Bevis Colon
2	CCO0065	Carson Caesar Olson	Carson.Caesar.Olson@dtaa.com	SoftwareEngineer	Carl Scott Hooper
1	HAC0651	Halee Autumn Craft	Halee.Autumn.Craft@dtaa.com	SoftwareEngineer	Gabriel Camden Joseph
4	PJE0714	Paki Jason Emerson	Paki.Jason.Emerson@dtaa.com	Salesman	Castor Chaim Combs
3	LCD0715	Lillith Chanda Delgado	Lillith.Chanda.Delgado@dtaa.com	Salesman	Castor Chaim Combs
1	GGG0825	Garrison George Guerrero	Garrison.George.Guerrero@dtaa.com	Salesman	Dennis Carson Mendoza
5	KHC0465	Kelly Hashim Carr	Kelly.Hashim.Carr@dtaa.com	Salesman	Bo Mira Dillard
5	CCD0463	Chancellor Cody Douglas	Chancellor.Cody.Douglas@dtaa.com	Salesman	Bo Mira Dillard
5	LKF0701	Levi Keegan Foley	Levi.Keegan.Foley@dtaa.com	Salesman	Castor Chaim Combs
6	DHB0696	Damon Hyatt Boyle	Damon.Hyatt.Boyle@dtaa.com	Salesman	Castor Chaim Combs
1	CLN1234	Chris L North	Chris.L.North@dtaa.com	Professor	Calvin J Ribbens
2	GHG0667	Giacomo Hamilton Giles	Giacomo.Hamilton.Giles@dtaa.com	ProductionLineWorker	Abraham Perry Reyes
6	SAO0920	Simone Amber Owens	Simone.Amber.Owens@dtaa.com	ProductionLineWorker	Theodore Upton Barry
2	SKC0670	Scarlett Kerry Charles	Scarlett.Kerry.Charles@dtaa.com	ProductionLineWorker	Abraham Perry Reyes
4	JAL0281	Justine Amy Lara	Justine.Amy.Lara@dtaa.com	ProductionLineWorker	Nadine Anjolie Gregory
1	RDC0283	Ruth Daphne Carver	Ruth.Daphne.Carver@dtaa.com	ProductionLineWorker	Nadine Anjolie Gregory
3	CSF0929	Chaney Sean Fuentes	Chaney.Sean.Fuentes@dtaa.com	ProductionLineWorker	Theodore Upton Barry
1	HRL0797	Hanae Rhiannon Lindsey	Hanae.Rhiannon.Lindsey@dtaa.com	ProductionLineWorker	Hector Davis Ellison
6	DBM0279	Dane Bert Mcfadden	Dane.Bert.Mcfadden@dtaa.com	ProductionLineWorker	Nadine Anjolie Gregory
5	JQV0922	Jada Quinn Vang	Jada.Quinn.Vang@dtaa.com	ProductionLineWorker	Theodore Upton Barry
5	DSL0441	Denise Serina Le	Denise.Serina.Le@dtaa.com	ProductionLineWorker	Brenna Martha Russell
2	IQW0294	Ima Quintessa Williams	Ima.Quintessa.Williams@dtaa.com	ProductionLineWorker	Nadine Anjolie Gregory
4	NPB0217	Noah Perry Bolton	Noah.Perry.Bolton@dtaa.com	Physicist	Evelyn Xena Santana
6	MBS0355	Megan Brynn Snyder	Megan.Brynn.Snyder@dtaa.com	Physicist	Haviva Jada Molina
6	HMW0274	Halee Morgan Wilkins	Halee.Morgan.Wilkins@dtaa.com	MechanicalEngineer	Melissa Ruby Knight
3	PAS0349	Palmer Ahmed Sutton	Palmer.Ahmed.Sutton@dtaa.com	Mathematician	Haviva Jada Molina
2	KAK0992	Kenyon Arthur Kent	Kenyon.Arthur.Kent@dtaa.com	Mathematician	Josephine Sylvia Parsons
2	WSP0210	Whilemina Shafira Perry	Whilemina.Shafira.Perry@dtaa.com	ITAdmin	Jaquelyn Yoshi Mckinney
5	CCH0959	Cedric Cyrus Harrison	Cedric.Cyrus.Harrison@dtaa.com	IndustrialEngineer	Desiree Claudia Booth
1	DZJ0261	Dominic Zeph Johns	Dominic.Zeph.Johns@dtaa.com	HardwareEngineer	Evan Bevis Colon
5	MJM0080	Martin Jonathan Martin	Martin.Jonathan.Martin@dtaa.com	ElectricalEngineer	Jolene Martha Blake
6	HMD0974	Hector Martin Daugherty	Hector.Martin.Daugherty@dtaa.com	ElectricalEngineer	Ann Hannah Dickerson
6	FKR0134	Francesca Kylie Russo	Francesca.Kylie.Russo@dtaa.com	AdministrativeAssistant	Suki Ginger Estes

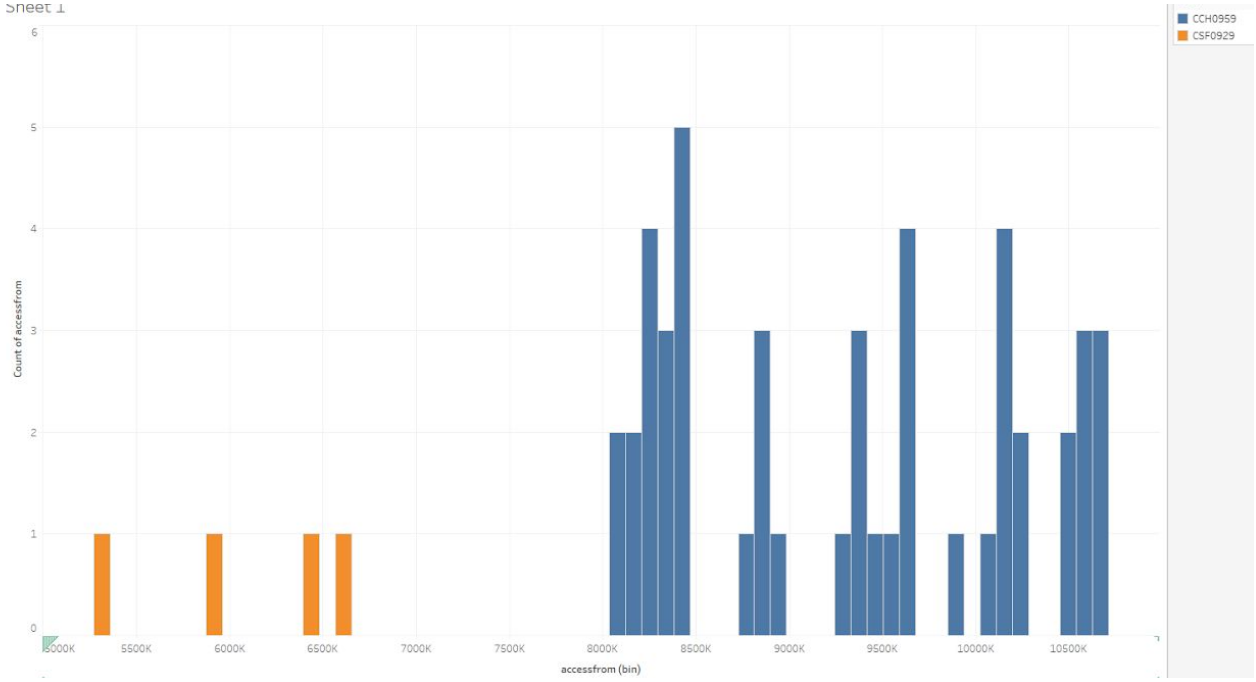
1.10 Filtering The Data By Fired Employees

We realized that we wanted to focus some more on the employees that were fired/quit their jobs. For such a purpose, it would be prudent to filter down the data to include just those rows that related to these employees. This would reduce the size of the data to be processed as well as visualized. For this, we made use of custom Python code that read in the list of IDs and emails belonging to the fired employees, and in the new set of files, excluded any row that didn't include at least one of these. This resulted in a significantly smaller data set to parse.

1.11 Device Usage of Fired Employees

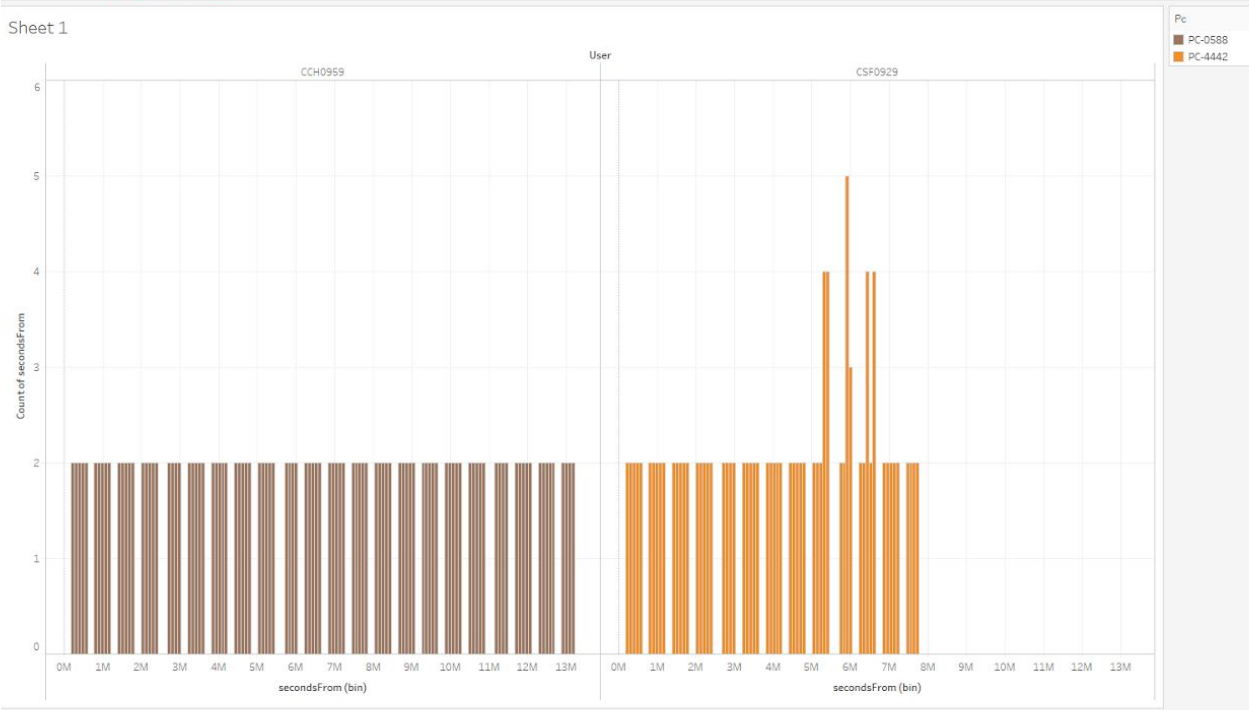
While plotting the device usage for employees across a timeline, we noticed a spike in activity for two employees, Chaney Sean Fuentes and Cedric Cyrus Harrison. The otherwise silent Chaney suddenly had a short burst of activity, and Cedric's device usage suddenly jumped up compared to normal in the days before he left the company.

1.12 Wikileaks and Lockheed Website Visitation



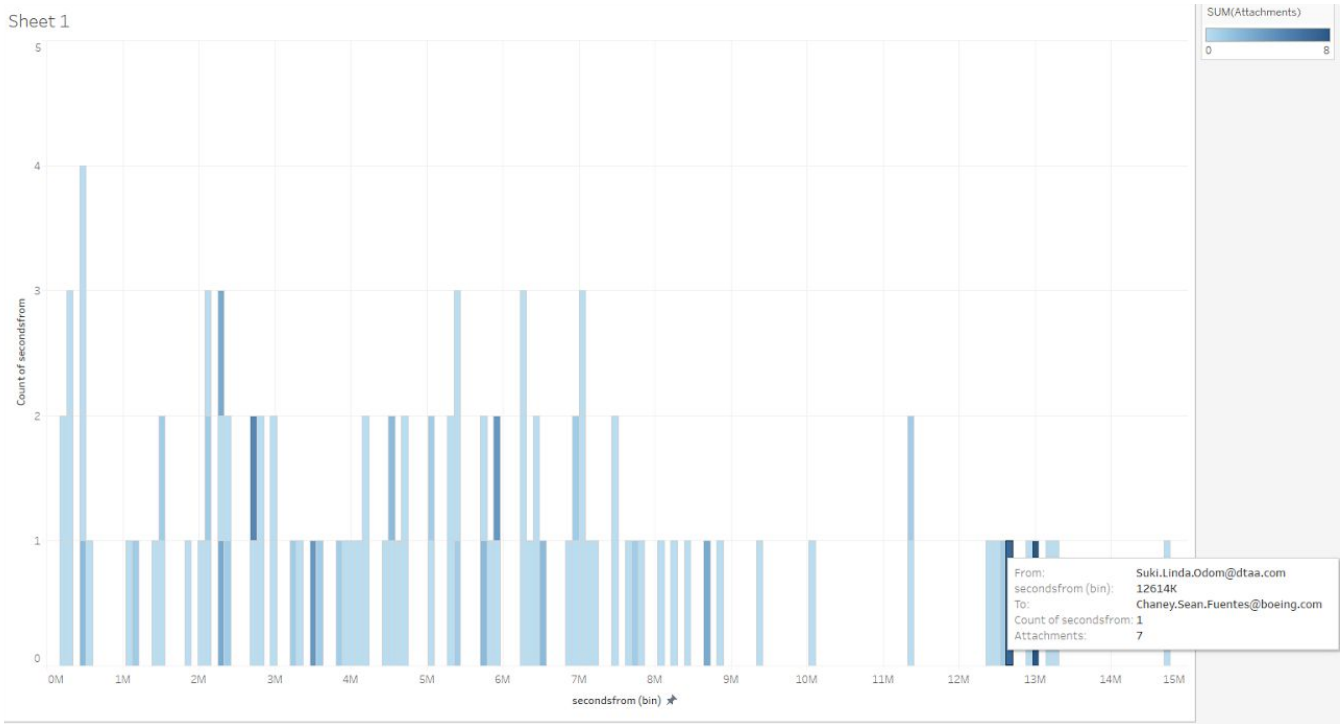
On further investigation of these two leads, we noticed unusual things about their website visitation. We observed that Chaney visited http://wikileaks.org on dates coinciding with their device usage spike, and Cedric visited http://lockheedmartinjobs.com multiple times just before leaving the company. These were two of the only websites in the data that didn't have the usual format.

1.13 Chaney and Cedric Logon Info



We then checked the logon info dataset to see if there were any similar spikes, and sure enough, there was a sharp spike in Chaney's logons and logoffs corresponding with the dates the devices usage spiked, in as well as dates wikileaks was visited.

1.14 Suki Emails to Chaney



On analyzing the email data of Chaney Sean Fuentes, we notice that even after Chaney had left the company, an employee named Suki Linda Odom was still sending them emails, some with attachments. What was especially noticeable was that after leaving the company, Chaney's email changes to being from the @boeing.com domain name.

1.15 Mail connections

After identifying that suspicious people is among employees that currently do not work in the company, we queried the emails among all those employes and generate a force directed graph that could show us the relations among former employees, and if they communicated to any of our current suspects. This is shown in Figure 5. The thickness of the lines is based on the number of emails sent and received between two nodes. In this Figure, we could notice that there is a group of employees who are place together in the middle of the graph and the edges connecting them are thick.

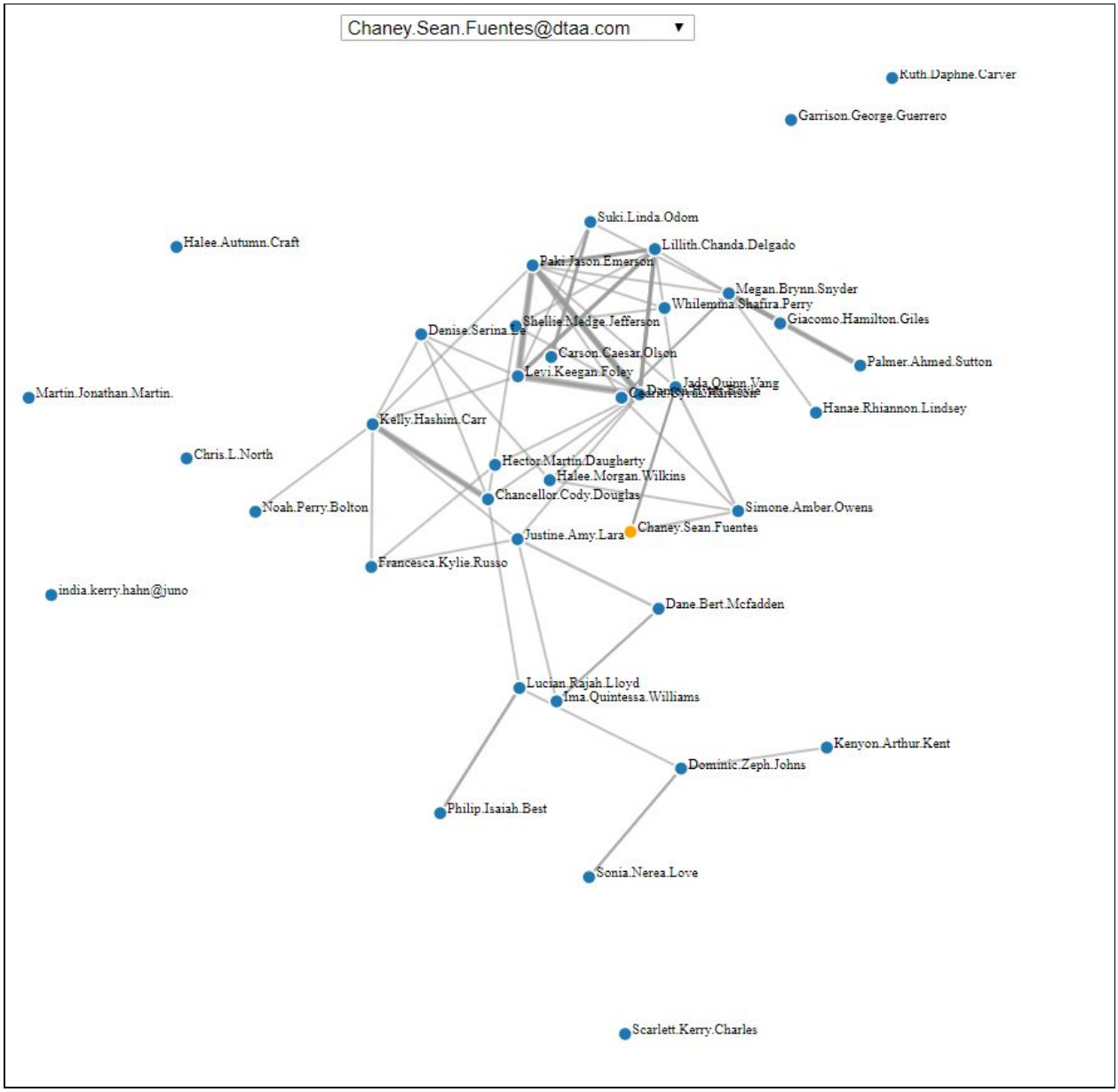


Figure 5. Force Directed Graph of Former employees based on their email interactions.

The suspicious employes are marked with orange nodes in the graph in Figure 6. We can see that the suspicious employees are part of a connected graph. That means that any information can be passed from one suspicious employee to another directly or through other employees.

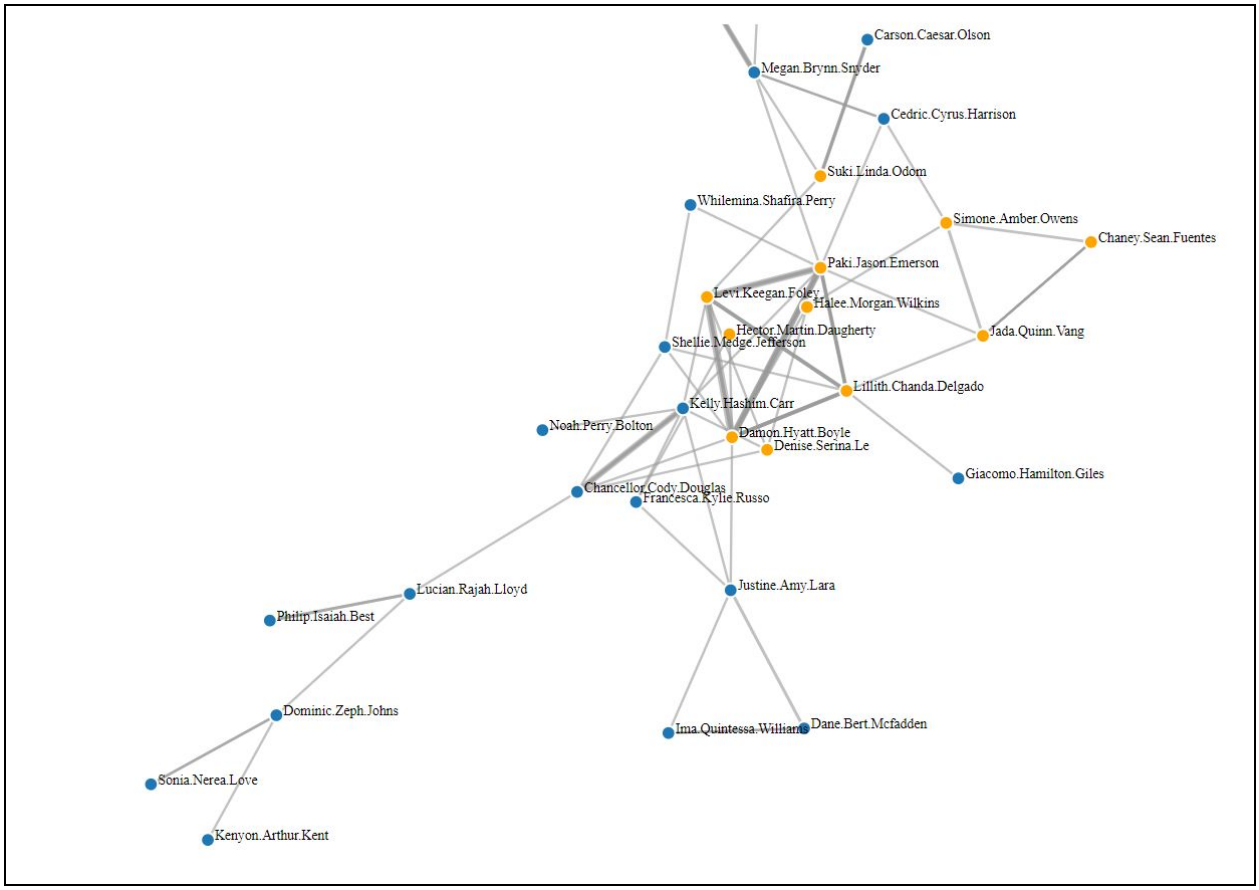


Figure 6. Suspicious employees in the graph of email connections

After, we found that there some employees who have email addresses with external domains. That is, emails with domains different @dtaa.com. Then we generate a graph of the emails sent to external domains from former employees to check if there was any leakage of information. In this graph (Figure 7), the blue nodes represent the former employees and the orange nodes represent external email addresses.

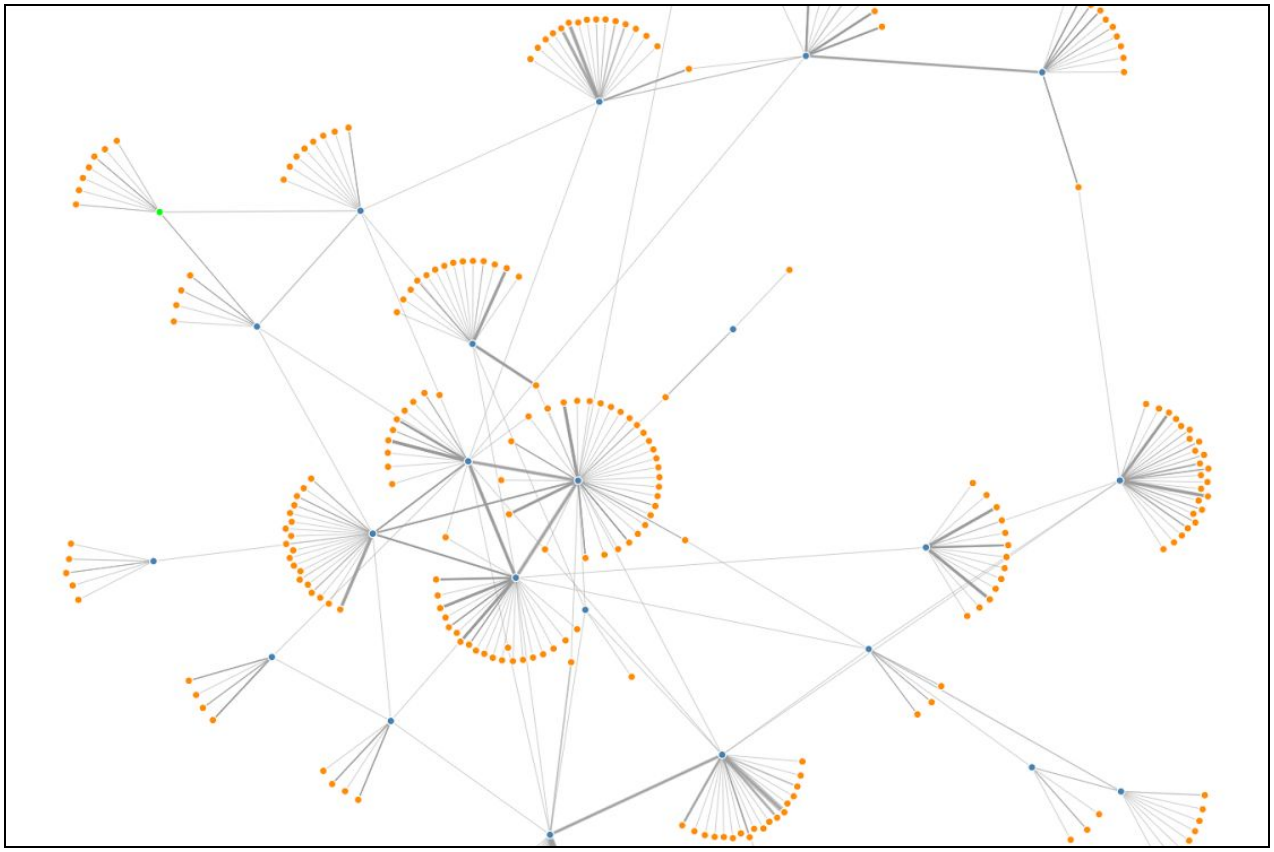


Figure 7. Email connection among former employees and external email addresses.

Looking at this graph, we found that there are four former employees who communicate between them and, based on line thickness, they have sent and receive many email between each other. In Figure 8, we zoomed in the area containing the mentioned employees.

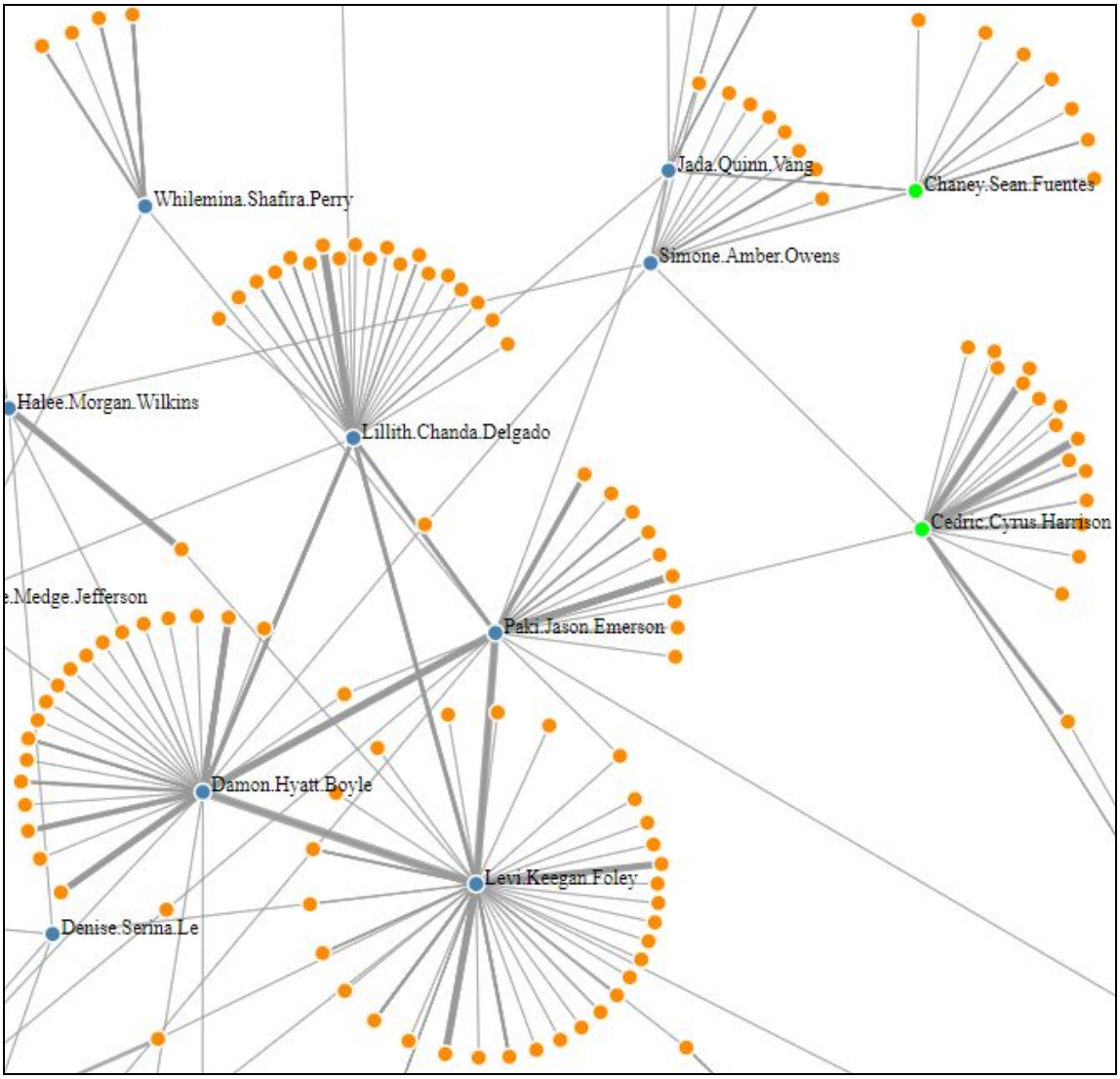


Figure 8. Former employees that are strongly connected

These employees are listed in the table below. In this table, we could see that the former employees are salesmen and have the same supervisor. So, the supervisor was also considered a suspect employee that is still in the company. In addition, among these employees, there is Levi Keegan Foley, who previously was detected as a suspicious person due to a consistently high device activity which suddenly stopped on a certain date. Moreover, these employees are connected to other suspicious employees like Cedric Cyrus Harrison and Chaney Sean Fuentes though the employes Jada Quinn Vang and Simeone Amber Owens. Cedric is a suspect since he has sent many emails including attachments to an email address with a domain of a possible competitor company as shown in Figure 9.

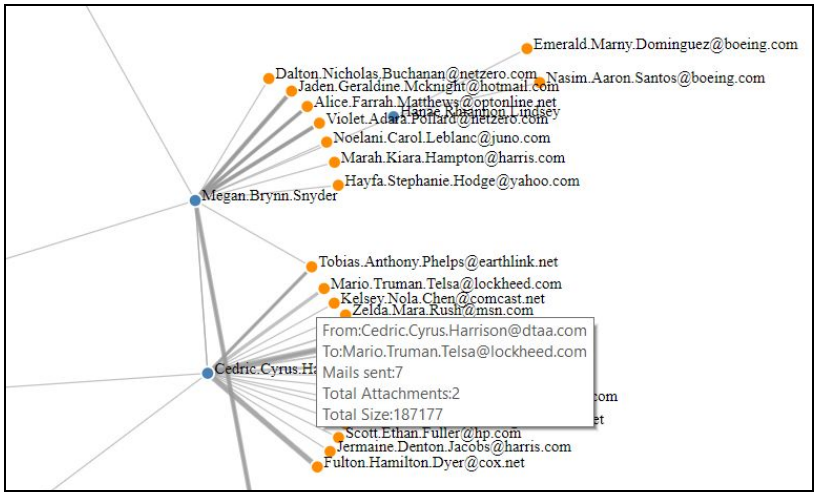


Figure 9. Leakage of information through email

Name	Id	Email	Role	Supervisor	Employee from-to
Damon Hyatt Boyle	DHB0696	Damon.Hyatt.Boyle@dtaa.com	Salesman	Castor Chaim Combs	May-Oct
Lillith Chanda Delgado	LCD0715	Lillith.Chanda.Delgado@dtaa.com	Salesman	Castor Chaim Combs	May-Jul
Levi Keegan Foley	LKF0701	Levi.Keegan.Foley@dtaa.com	Salesman	Castor Chaim Combs	May-Sep
Paki Jason Emerson	PJE0714	Paki.Jason.Emerson@dtaa.com	Salesman	Castor Chaim Combs	May-Aug

1.16 Fired Employee Supervisors

Since our suspicious employees were among the fired employees, we initially used the Python code and added a new column to see each of their supervisors, to see if there was any correlation.

Fired Employees:

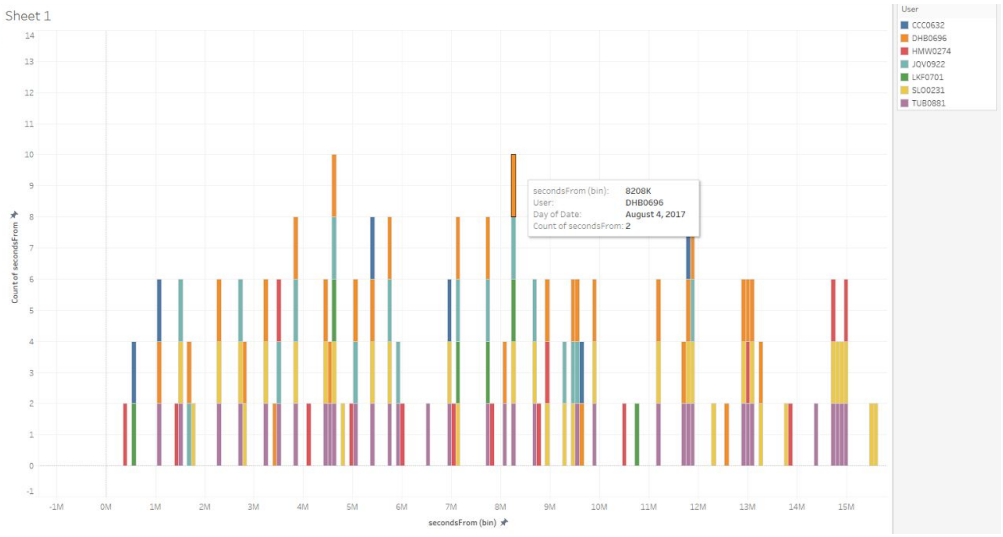
Months Worked	Last Month	Name	Supervisor
6	10/01/2017	Simone Amber Owens	Theodore Upton Barry
6	10/01/2017	Megan Brynn Snyder	Haviva Jada Molina
2	06/01/2017	Scarlett Kerry Charles	Abraham Perry Reyes
6	10/01/2017	Damon Hyatt Boyle	Castor Chaim Combs
1	05/01/2017	Ruth Daphne Carver	Nadine Anjolie Gregory
3	07/01/2017	Palmer Ahmed Sutton	Haviva Jada Molina
3	07/01/2017	Chaney Sean Fuentes	Theodore Upton Barry
2	06/01/2017	Giacomo Hamilton Giles	Abraham Perry Reyes
6	10/01/2017	Dane Bert Mcfadden	Nadine Anjolie Gregory
6	10/01/2017	Halee Morgan Wilkins	Melissa Ruby Knight
5	09/01/2017	Jada Quinn Vang	Theodore Upton Barry
4	08/01/2017	Paki Jason Emerson	Castor Chaim Combs
6	10/01/2017	Hector Martin Daugherty	Ann Hannah Dickerson
3	07/01/2017	Lucian Rajah Lloyd	Hunter Stephen Slater
5	09/01/2017	Levi Keegan Foley	Castor Chaim Combs
1	05/01/2017	Sonia Nerea Love	Evan Bevis Colon
2	06/01/2017	Ima Quintessa Williams	Nadine Anjolie Gregory
1	05/01/2017	Garrison George Guerrero	Dennis Carson Mendoza
2	06/01/2017	Whilemina Shafira Perry	Jaquelyn Yoshi Mckinney
4	08/01/2017	Justine Amy Lara	Nadine Anjolie Gregory
1	05/01/2017	Chris L North	Calvin J Ribbens
2	06/01/2017	Carson Caesar Olson	Carl Scott Hooper
6	10/01/2017	Shellie Medge Jefferson	William Vernon Booth
1	05/01/2017	Hanae Rhiannon Lindsey	Hector Davis Ellison
4	08/01/2017	Noah Perry Bolton	Evelyn Xena Santana
5	09/01/2017	Chancellor Cody Douglas	Bo Mira Dillard
5	09/01/2017	Martin Jonathan Martin	Jolene Martha Blake
5	09/01/2017	Denise Serina Le	Brenna Martha Russell
3	07/01/2017	Lillith Chanda Delgado	Castor Chaim Combs
1	05/01/2017	Dominic Zeph Johns	Evan Bevis Colon
2	06/01/2017	Kenyon Arthur Kent	Josephine Sylvia Parsons
5	09/01/2017	Cedric Cyrus Harrison	Desiree Claudia Booth
6	10/01/2017	Francesca Kylie Russo	Suki Ginger Estes
1	05/01/2017	Halee Autumn Craft	Gabriel Camden Joseph
5	09/01/2017	Kelly Hashim Carr	Bo Mira Dillard
2	06/01/2017	Philip Isaiah Best	Hunter Stephen Slater

We noticed that a few names popped up often, and so decided to look for groupings that could shed further light on what was going on, and who else might be involved.

We then presented the following table, which groups the fired employees who have the the same supervisors.

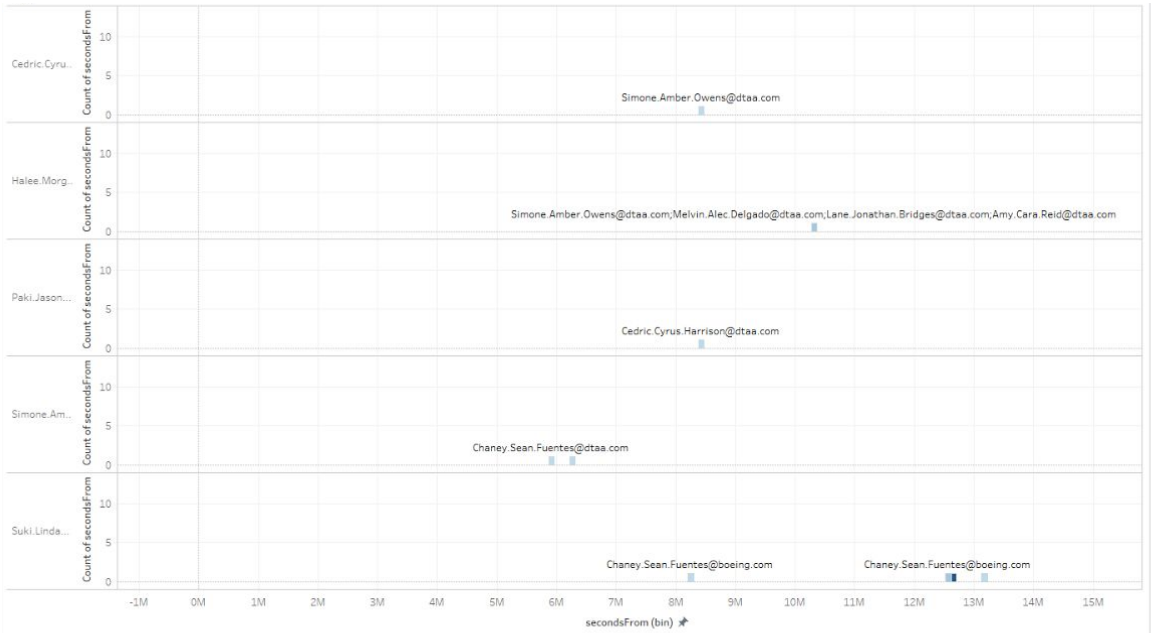
F6	F2
Calvin J Ribbens	Chris L North
Dennis Carson Mendoza	Garrison George Guerrero
Gabriel Camden Joseph	Halee Autumn Craft
Hector Davis Ellison	Hanae Rhiannon Lindsey
Carl Scott Hooper	Carson Caesar Olson
Evan Bevis Colon	Dominic Zeph Johns
	Sonia Nerea Love
Jaquelyn Yoshi Mckinney	Whilemina Shafira Perry
Josephine Sylvia Parsons	Kenyon Arthur Kent
Abraham Perry Reyes	Giacomo Hamilton Giles
	Scarlett Kerry Charles
Evelyn Xena Santana	Noah Perry Bolton
Brenna Martha Russell	Denise Serina Le
Desiree Claudia Booth	Cedric Cyrus Harrison
Hunter Stephen Slater	Lucian Rajah Lloyd
	Philip Isaiah Best
Jolene Martha Blake	Martin Jonathan Martin
Ann Hannah Dickerson	Hector Martin Daugherty
Melissa Ruby Knight	Halee Morgan Wilkins
Suki Ginger Estes	Francesca Kylie Russo
William Vernon Booth	Shellie Medge Jefferson
Haviva Jada Molina	Megan Brynn Snyder
	Palmer Ahmed Sutton
Bo Mira Dillard	Chancellor Cody Douglas
	Kelly Hashim Carr
Nadine Anjolie Gregory	Dane Bert Mcfadden
	Ima Quintessa Williams
	Justine Amy Lara
	Ruth Daphne Carver
Theodore Upton Barry	Chaney Sean Fuentes
	Jada Quinn Vang
	Simone Amber Owens
Castor Chaim Combs	Damon Hyatt Boyle
	Levi Keegan Foley
	Lillith Chanda Delgado
	Paki Jason Emerson

1.17 Shared PC Access by Suspects



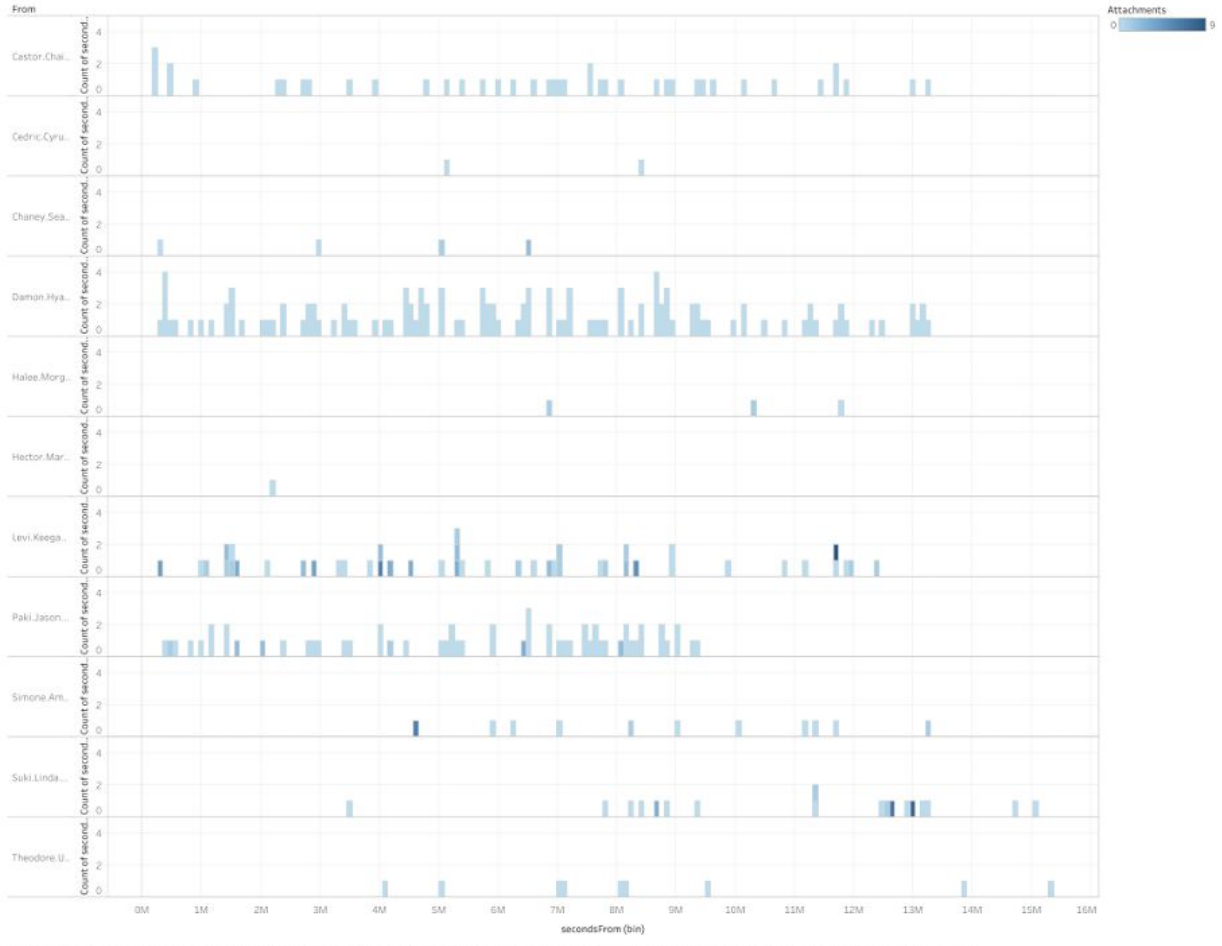
After examining histograms of our suspects' logon activity, we noticed that several suspects, in addition to logging into their primary computer, logged into the shared computer PC-7165. We explored further by looking at the logon activity among our suspects on the shared computer and found two spikes in activity on June 23rd and August 4th.

1.18 Temporal Emails between Suspects



When trying to piece together the transfer of information between suspects, we attempted to look at the temporal trends in emails between our suspects which was not possible with our

network visualization shown in Section 1.15. We used Tableau and a juxtaposed layout of histograms where each view shows emails sent from a single employee. After tracing the flow of emails on a reduced set of suspects shown in the figure above, we moved onto the full set as shown in the figure below. Most notably, the fired employee Simone Amber Owens had contact with Chaney Sean Fuentes on the same days Chaney visited Wikileaks.



1.19 Device Info and Logon Info of Final Suspects

As a summary of our final list of suspects, we produced histograms of device usage and logon activity and scanned over both sets of data as a single multiple-view visualization consisting of two facets, device activity and logon activity, each with juxtaposed histograms of individual employees. All of the data was visualized with respect to an aligned X axis.

2. Hypotheses

From the analysis methods listed above, we identified a few people that seemed suspicious.

id	date	email_to	email_from	size	attachments
{V8M5-W3UP43OI-3551CAZW}	6/2/2017 15:52	Palmer.Nathaniel.Church@dtaa.com	Chris.L.North@vt.edu	46845	0

month	employee_name	user_id	email	role	supervisor
7/1/2017	Chaney Sean Fuentes	CSF0929	Chaney.Sean.Fuentes@dtaa.com	Production Line Worker	Theodore Upton Barry
9/1/2017	Cedric Cyrus Harrison	CCH0959	Cedric.Cyrus.Harrison@dtaa.com	Industrial Engineer	Desiree Claudia Booth
5/1/2017	Palmer Nathaniel Church	PNC0141	Palmer.Nathaniel.Church@dtaa.com	Manager	Buffy Meredith Giles
5/1/2017	Chris L North	CLN1234	Chris.L.North@dtaa.com	Professor	Calvin J Ribbens
11/1/2017	Suki Linda Odom	SLO0231	Suki.Linda.Odom@dtaa.com	Computer Programmer	Carl Scott Hooper
9/1/2017	Levi Keegan Foley	LKF0701	Levi.Keegan.Foley@dtaa.com	Salesman	Castor Chaim Combs
10/1/2017	Damon Hyatt Boyle	DHB0696	Damon.Hyatt.Boyle@dtaa.com	Salesman,	Castor Chaim Combs
7/1/2017	Lillith Chanda Delgado	LCD0715	Lillith.Chanda.Delgado@dtaa.com	Salesman,	Castor Chaim Combs

8/1/2017	Paki Jason Emerson	PJE0714	Paki.Jason.Emerson@dtaa.com	Salesman,	Castor Chaim Combs
10/1/2017	Halee Morgan Wilkins	HMW0274	Halee.Morgan.Wilkins@dtaa.com	Mechanical Engineer	Melissa Ruby Knight
9/1/2017	Jada Quinn Vang	JQV0922	Jada.Quinn.Vang@dtaa.com	Production Line Worker	Theodore Upton Barry
10/1/2017	Simone Amber Owens	SAO0920	Simone.Amber.Owens@dtaa.com	Production Line Worker	Theodore Upton Barry
11/1/2017	Castor Chaim Combs	CCC0632	Castor.Chaim.Combs@dtaa.com	Manager	Perry Reese Morton
11/1/2017	Theodore Upton Barry	TUB0881	Theodore.Upton.Barry@dtaa.com	Assembly Supervisor	Gisela Jeanette Watkins
9/1/2017	Denise Serina Le	DSL0441	Denise.Serina.Le@dtaa.com	Production Line Worker	Brenna Martha Russell
10/1/2017	Hector Martin Daugherty	HMD0974	Hector.Martin.Daugherty@dtaa.com	Electrical Engineer	Ann Hannah Dickerson

Chaney Sean Fuentes became one of the lead suspects after it was found that Chaney had accessed http://wikileaks.com. Apart from it being one of the only valid sites in the dataset, these accesses coincided with spikes in device and logon activity. The spikes in device info in particular are very telling, because apart from those on and around the dates where wikileaks was accessed, Chaney has no other device accesses in his entire time at the company.
The dates when this suspicious activity took place wer **July 1st, July 8th, July 14th, and July 16th.**

Cedric Cyrus Harrison accessed http://lockheedmartinjobs.com, one of the few websites that led to an actual working website. DTAA is a defence technology company, possibly a competitor to/client of Lockheed Martin. Cedric also sent 7 emails to Mario.Truman.Telsa@lockheed.com

during the month of August. Cedric also had a spike in device usage and logon activity right before his leaving the company.

Suki Linda Odom is the only employee that kept emailing Chaney Sean Fuentes even after he left the company. These emails extended to even after he had a new email at Boeing and contained many attachments.

Chris L North is the only person with the role of 'Professor' in the organization, and left after a month. Chris L North's supervisor, Calvin J Ribbens, does not exist in the database. A month after Chris left the company, he shows up in an email from a different email address.
We noticed that in the email_info.csv file, there was an email sent from 'Chris.L.North@vt.edu', an email address with a domain that didn't belong to the company. A search of employee_info.csv also did not show that email address. Searching for Chris L North in employee_info.csv, however, did find a record of an employee by that name who was only with the company for a month, and had the unique role of 'Professor', with a supervisor whose name did not exist anywhere else in the data.

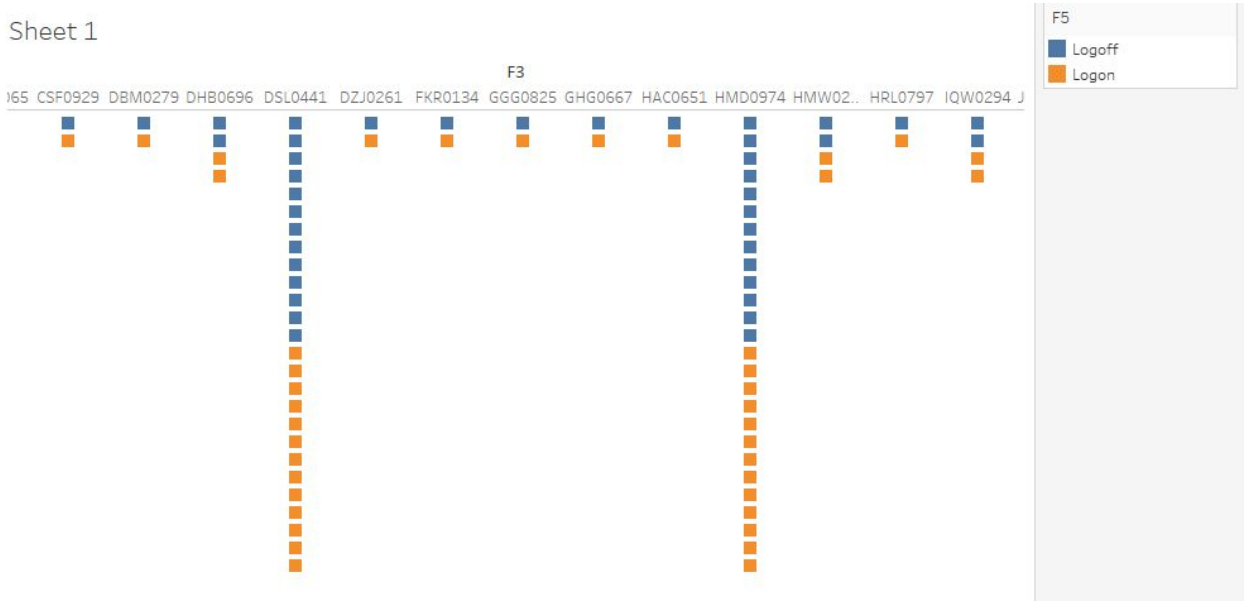
Palmer Nathaniel Church, who is a Manager working directly under the Vice President, is the only person who received an email from Chris L North's VT email, after Chris was no longer with the company.
This information also came from analyzing both email_info.csv and employee_info.csv together.

Levi Keegan Foley first came to our attention from our frequency analysis of device_info.csv presented in section 1.5. The Gantt chart we produced showed him as the only employee with consistently high device activity which suddenly stopped on a certain date. Our next step in investigating this lead was looking into the employee_info file for instances of his user id. There, we found his associated employee information and also discovered that he was fired at the time the histogram of this device activity showed his activity stopped. We also thought it was curious that he was a salesman due to his high device activity. We then looked up the titles of the other users with consistently high device activity and found he was the only salesman. From our analysis of fired employees, we discovered that very few other fired employees had much device activity or even appeared in device_info.csv at all. Lastly, we queried email_info for any instances of his email address and found that he sent an email just before he was fired with 6 attachments. We examined the distribution of number of attachments over all sent emails and found only a few had as many as 6 attachments.

Levi Keegan Foley, **Damon Hyatt Boyle**, **Lillith Chanda Delgado**, and **Paki Jason Emerson**, were all under the same supervisor, all parted ways with the company within a month of each other, and they all had a flurry of emails between them, once of the highest in the entire data set.

Jada Quinn Vang and Simone Amber Heard were fired production line workers under the same supervisor as Cedric. Both Cedric and Chaney communicated with Simone, and Jada Quinn communicate with the group of fired salesman above.

Denise Serina Le and **Hector Martin Daugherty** both logged onto multiple PCs despite their position in the company seemingly not requiring them to. Denise is a Production Line Worker and Hector is an Electrical Engineer. The spike in their PC usage can be seen here:



3. Attribution

While the visualizations and custom tools we used in our analysis were produced individually, lots of trends, patterns, and investigations of leads happened collaboratively during our team meetings. For example, much of the work in producing and verifying our list of suspects was a collaborative effort.

Christopher Wakeley:

- Scribed meetings and recorded activity
- Contributed to summary statistics
- Analyzed frequency of events and eventually produced the Gantt chart presented in section 1.5
- Continued analysis of frequency by producing histograms of logon activity presented in section 1.8
- Identified Levi Keegan Foley as a person of interest
- Produced many of the histograms in Tableau from sections 1.19, 1.18, 1.17, 1.14, 1.13, 1.12, and 1.11
- Contributed to identifying Chaney and Cedric as suspects based on device histograms
- Contributed to identifying common supervisors amongst fired employees.
- Helped analyze the force-directed graph of emails among suspects for trends.

Patricio Moreno:

- Create a database and load the data in Mysql to ease the queries.
- Analyze the logon dataset to find out suspicious PCs (Section 1.4).
- Combine email dataset with employee dataset to get more information about senders.
- Discover that there is a suspicious email sent with an email address which not correspond to any employee (Chris.L.North@vt.edu).
- Analyze frequency of visited websites grouping them by their domain (Section 1.6)
- Use sql queries to generate a JSON file containing the former employees information and the links between them.
- Create a force directed Graph (using D3.js) based on the email information included in JSON files to see the connections among suspicious users.
- Extend the JSON file to include external mail addresses
- Extend the graph to include connections to email addresses that do not belong to the company. It was done to check whether there was leakage of information through the email

Sourabh Shetty:

- Wrote a tool to:
    a. Convert employee_info.csv into a tree structure to easily identify direct relationships between employees.
    b. Print the entire employee hierarchy for each month.
    c. Find a user, their superiors, and their subordinates.
    d. Find all the employees that for any reason did not work during all recorded months of employment, i.e, either quit or were fired. Updated the logic to determine whether an employee counted as fired/quit or not.
    e. Added code to also list the Supervisor of the fired employees.
- Wrote another tool to read in a list of IDs and emails and used it to filter the data set to a subset that included only rows that featured at least one of the IDs/emails.
- Identified the suspicious line in employee_info with Calvin J Ribbens listed as a superior when that person does not exist in the rest of the database. That line also had Chris L North with a different email address than the one in the email_info.csv file.
- Contributed to identifying Chaney's wikileaks.org access.

- Identified 7 emails sent by Cedric Harrison in the month of April to an email under the @lockheedmartin.com domain name.
- Used Tableau to map logons and logoffs and noticed that Denise Serina Le and Hector Martin Daugherty were using a lot of different computers, and were the only employees that had such a variance in computer usage.
- Responsible for frequent manual lookup in all the csv data files.
- Attempted to use other tools like Gephi to find additional information.
- Kept the suspect table updated.