

R Shiny Investigative Tool User Guide

Exploratory Data Analysis

Locations for Exploration	2
Transactions for Exploration	3
Individual Card Transactions	5
Employees of GASTech	7
Email Correspondence	8

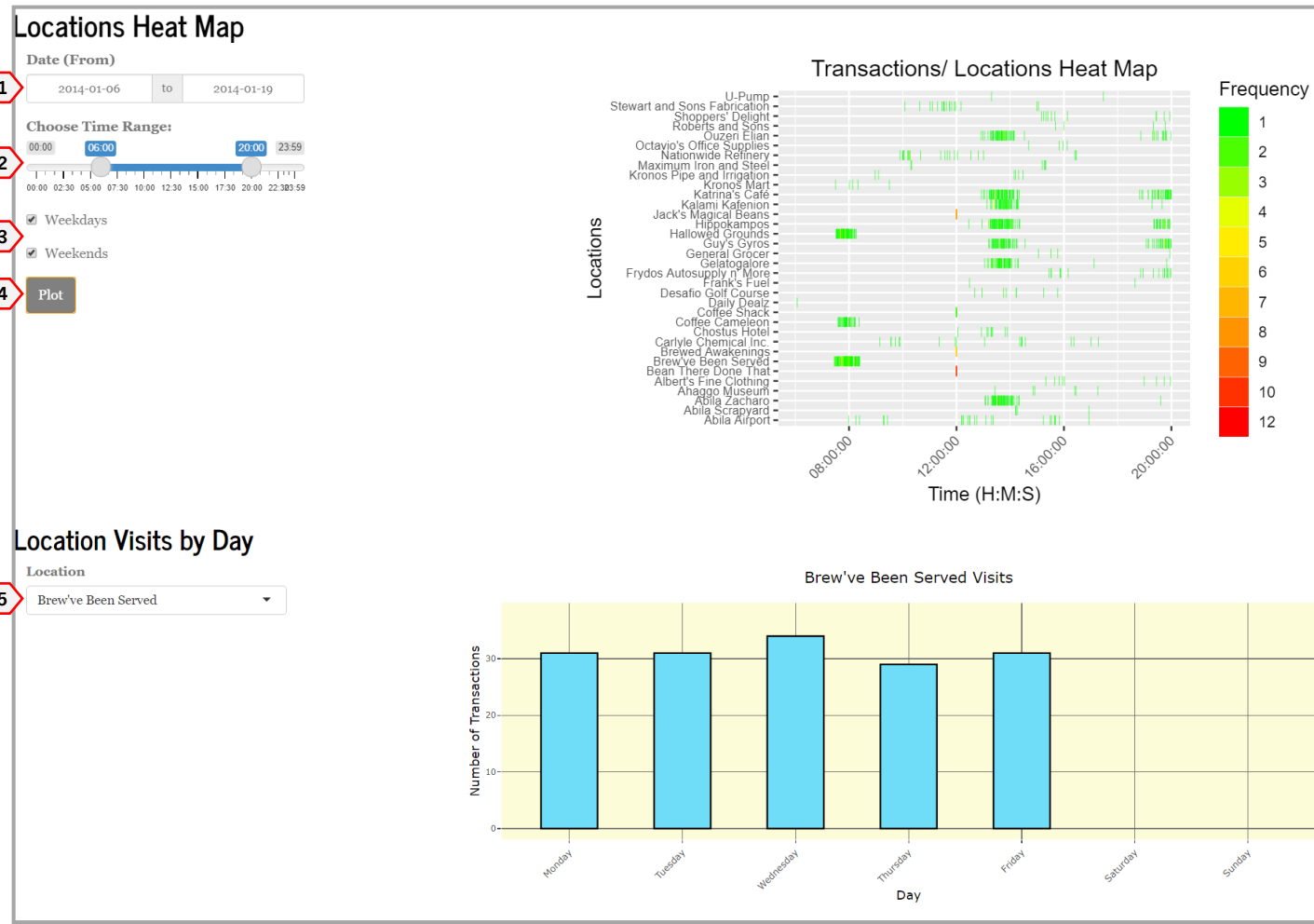
Inferential Statistics

Email Network Analysis	9
Networks	10
Employee Movement Plot	14
Transaction Amount Analysis	15

Locations for Exploration :: EDA

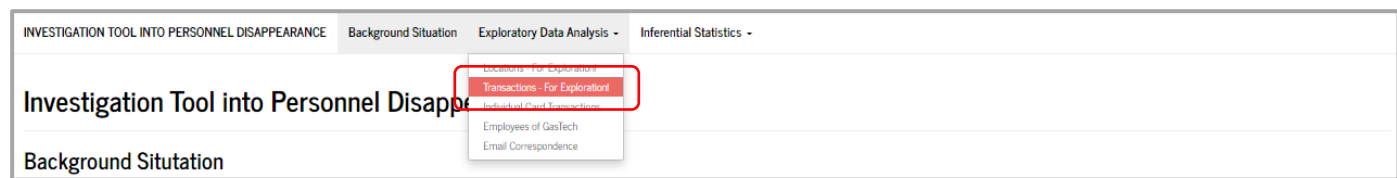


Under the tab **Exploratory Data Analysis**, select **Locations – For Exploration!**



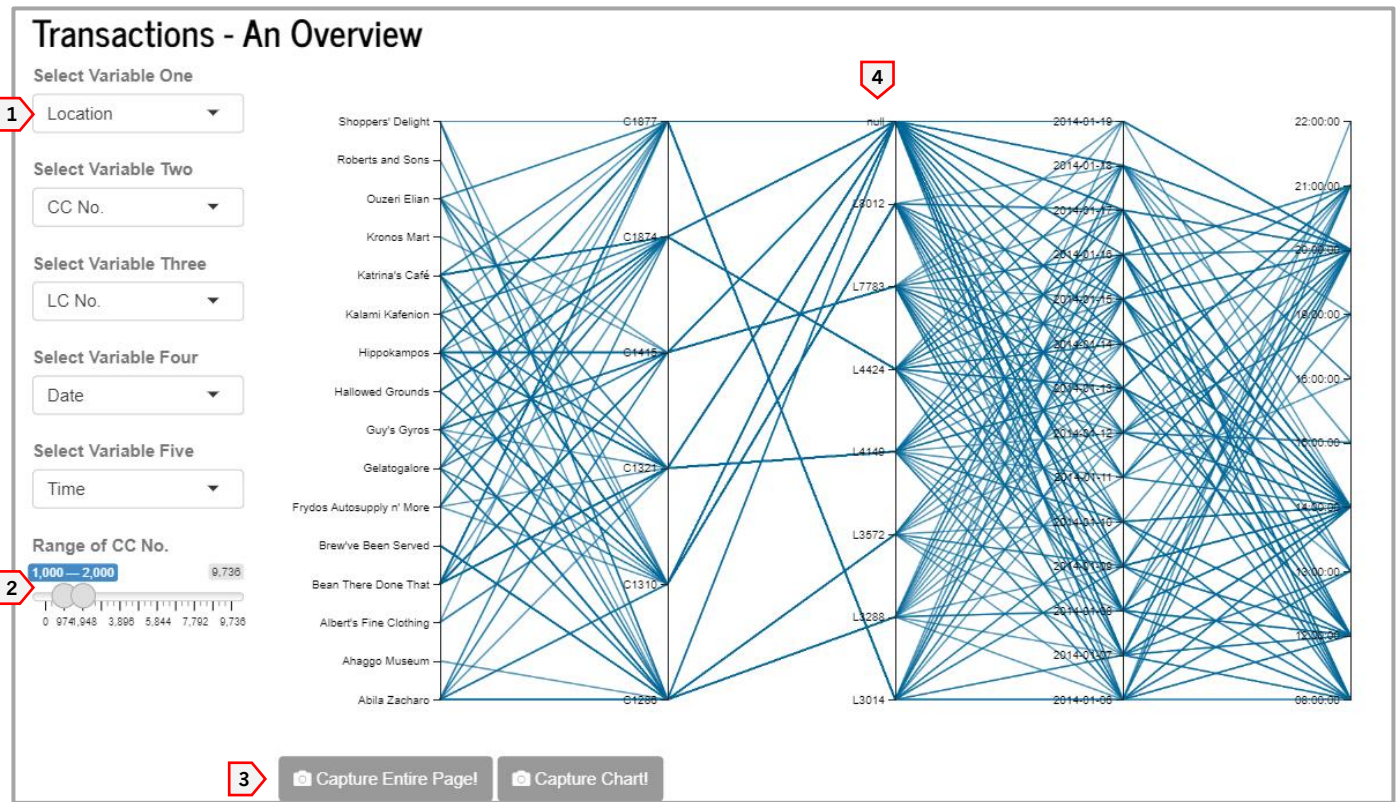
- [1] Select the range of dates for all the transactions.
- [2] Select the time range for the day to explore.
- [3] Could toggle between Weekdays, Weekends or both.
- [4] Push the “Plot” button and the heatmap would appear.
- [5] Select the location to explore, and the transactions for each day would be plotted on the right, inferring their operating days of the week.

Transactions for Exploration :: EDA

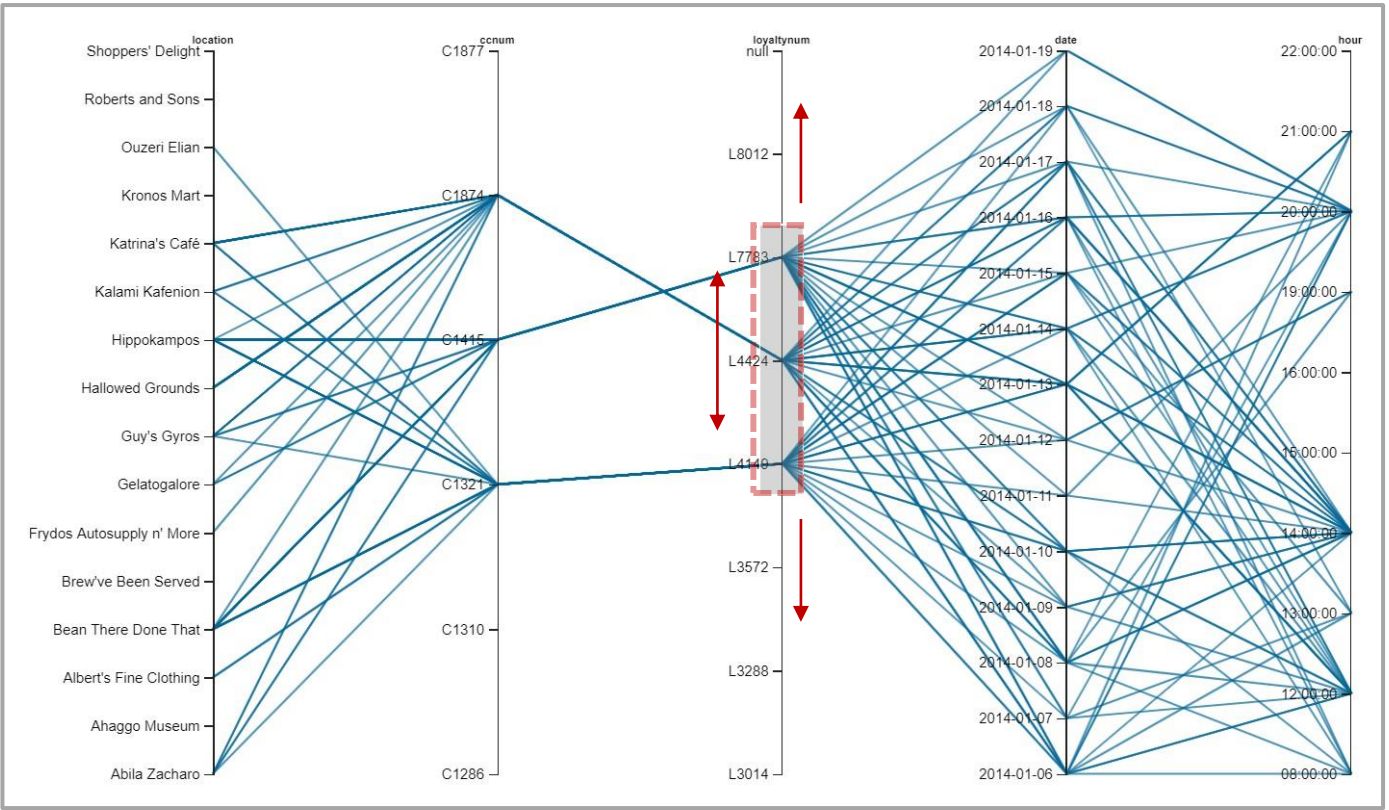


Under the tab **Exploratory Data Analysis**, select **Transactions – For Exploration!**

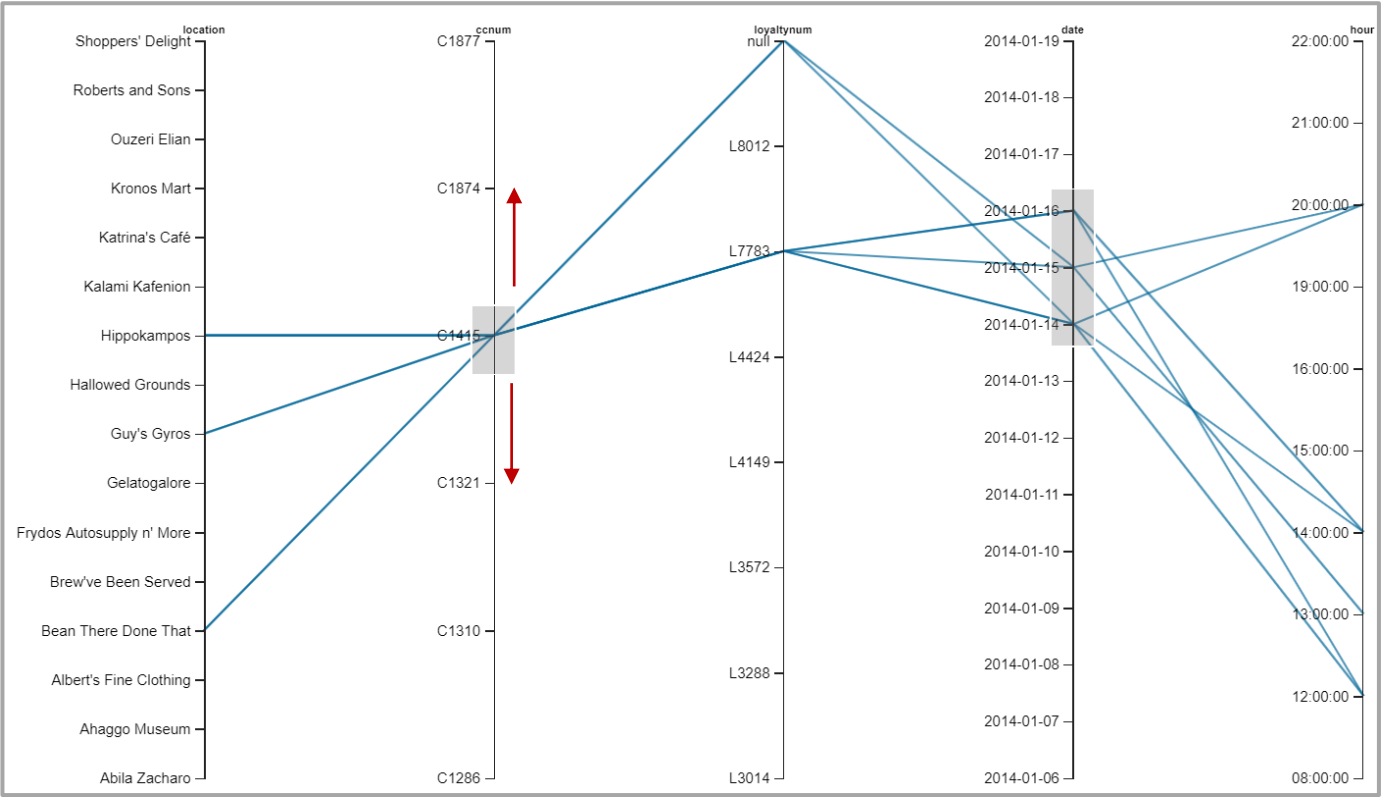
Parallel Coordinate Plots are useful charts for visualizing multivariate numerical data. They are ideal for comparing multiple variables at one go and to display the relationships between them. You will notice that each variable is given its own axis, and all axes are placed parallel to each other. Each observation is represented by a series of lines which connect across corresponding values across all axes.



- [1] Select the Variables to visualize the relationships for. Variable One would be placed on the leftmost axis, Variable Five on the rightmost axis.
- [2] Drag the slider to select the range of credit card numbers you would like the chart to display. There are 54 unique credit card numbers. Viewing by filtered ranges would allow you to view the lines, and thus the relationships, more clearly.
- [3] Screenshot buttons are provided for you to save your plots conveniently with one click, be it to compare later or include in a report.
- [4] The output plot is interactive. Hover over the plot till a crosshair appears, draw a rectangle over the range of variables you would like to see all connected lines for. You may also slide this grey box up and down the axis, connected lines (together with all their associations) which are within the range of this grey box would appear. Note that you can draw multiple boxes across variables and boxes are also resizable.



For e.g. you may wish to see which credit cards have transactions for the date period of 12th – 14th Jan -> Draw one grey box over 12th and 14th Jan, and another one over one credit card number. Slide the grey box on the credit card number axis and see all the connections appear and change.



Individual Card Transactions :: EDA

INVESTIGATION TOOL INTO PERSONNEL DISAPPEARANCE

Background Situation

Exploratory Data Analysis

Inferential Statistics

Investigation Tool into Personnel Disappearance

Locations - For Exploration

Transactions - For Exploration

Individual Card Transactions

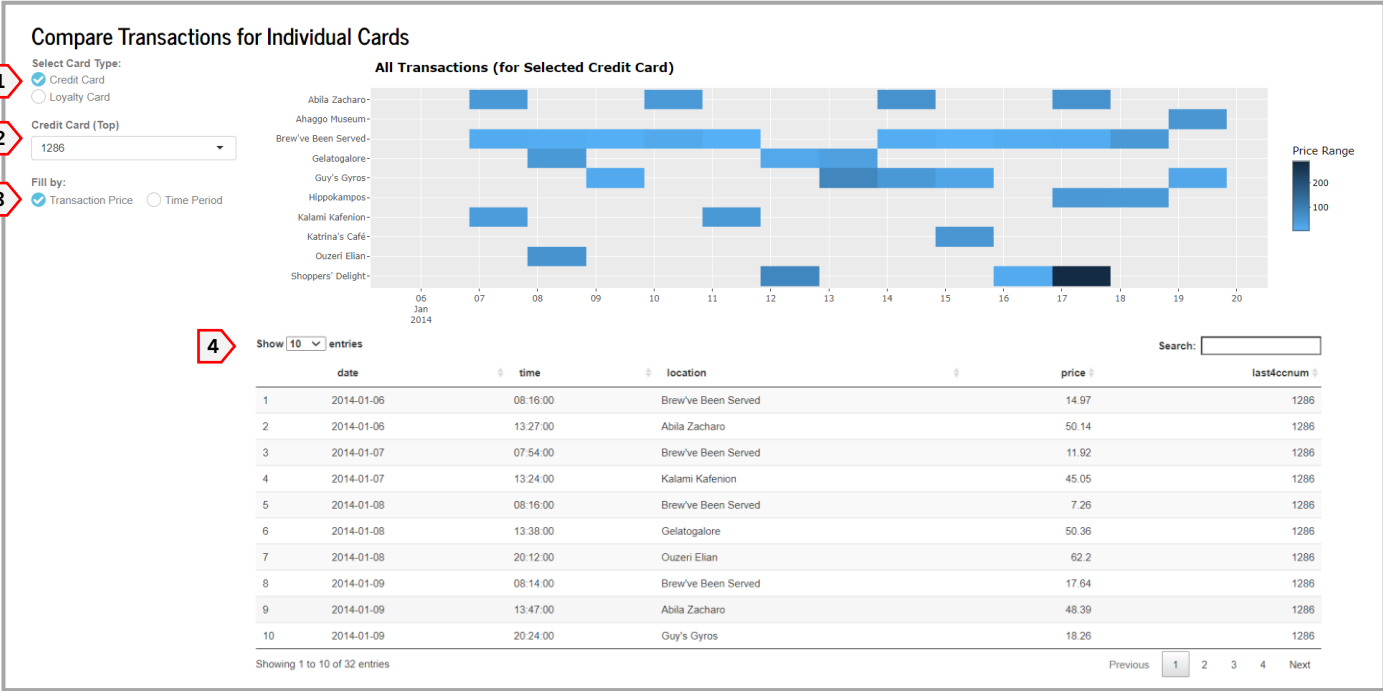
Employee's Contact

Email Correspondence

Background Situation

Under the tab **Exploratory Data Analysis**, select **Individual Card Transactions**

The tile plot shows the occurrences of transactions at each location on each day over the two weeks. In this site, two workspaces are provided for users to select two different card types for comparison against each other. One plot (paired with its own data table) is at the top, one at the bottom. Hover over each tile to see the details of that transaction.



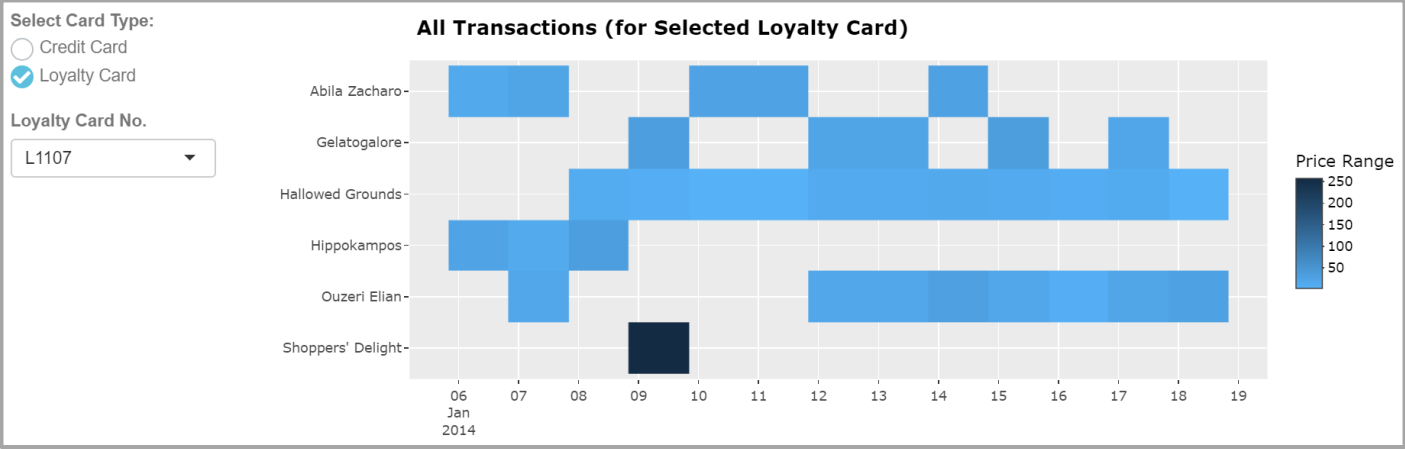
- [1] Select the type of card to view the transactions for.
- [2] The list of specific card numbers will appear in the dropdown list, depending on the type of card chosen.

[3] If credit card is chosen, you also get to choose whether to fill the tiles according to the magnitude of their transaction amounts, or the period of day the transaction took place at. The periods are defined as follows:

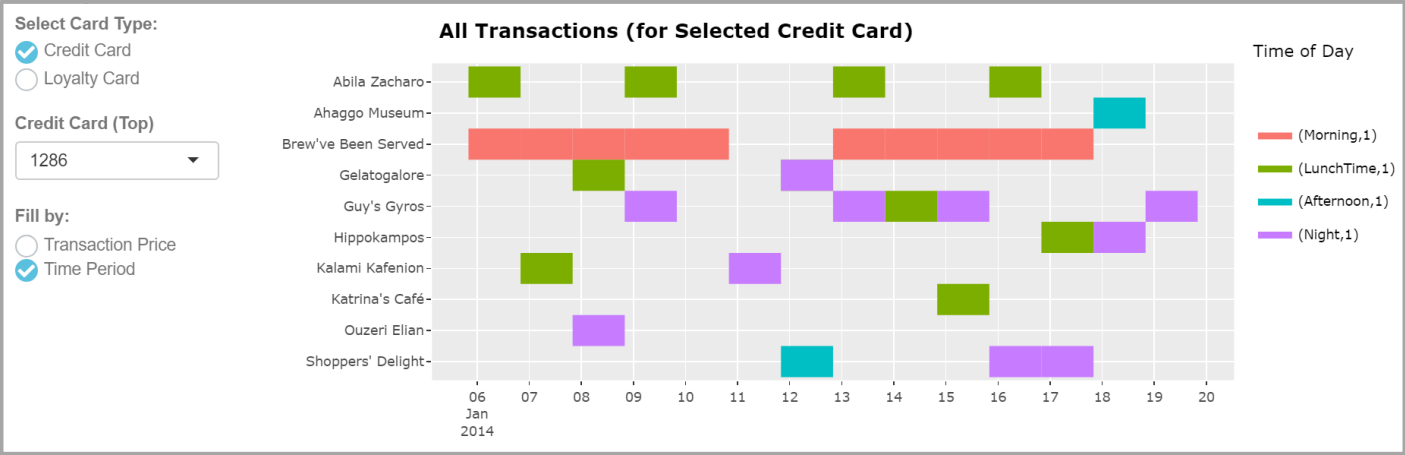
- Morning : 0701 – 1000 hrs
- Pre-Lunch : 1001 – 1200 hrs
- LunchTime : 1201 – 1430 hrs
- Afternoon : 1431 – 1700 hrs
- Evening : 1701 – 1900 hrs
- Night : 1901 – 0000 hrs
- Midnight : 0001 – 0700 hrs

[4] All transactions for the selected card will appear in the data table for further reference and analysis. Filtering and sorting can be done in the data table.

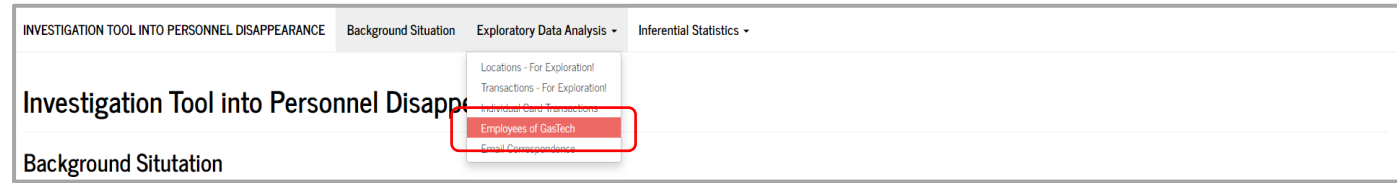
Plot for Loyalty Card



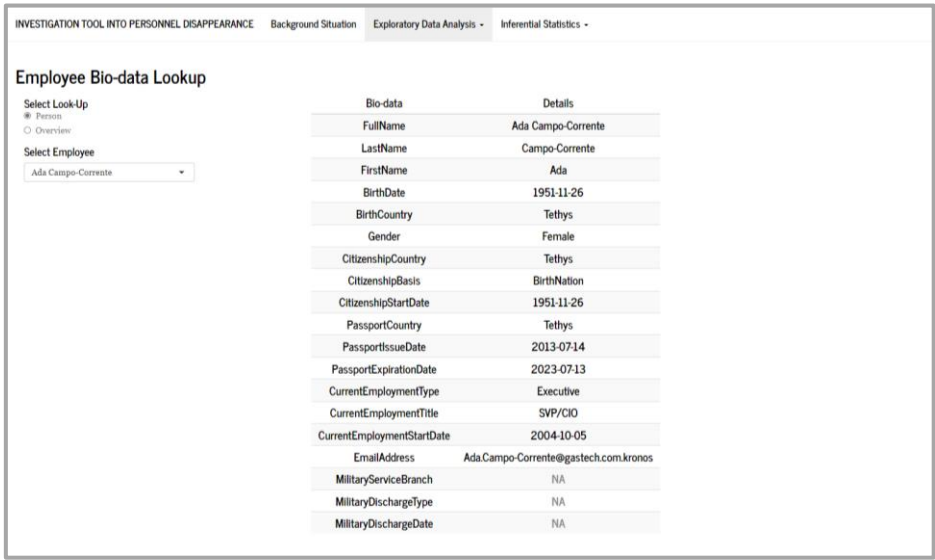
Plot for Credit Card, Fill by Time Period



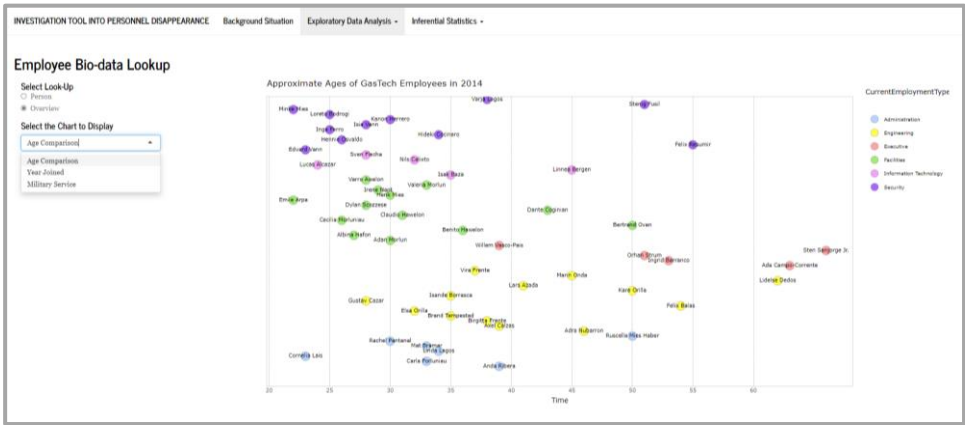
Employees of GASTech::: EDA



Under the tab **Exploratory Data Analysis**, select **Employees of GasTech**



- [1] Select the Person button
- [2] Select the Employee from the drop-down menu to view their biodata page



- [3] Select the Overview button
- [4] Select one of 3 charts that display the overall biodata of all GasTech Employees

Email Correspondence:: EDA

INVESTIGATION TOOL INTO PERSONNEL DISAPPEARANCE

Background Situation

Exploratory Data Analysis

Inferential Statistics

Investigation Tool into Personnel Disappearance

Background Situation

Locations - For Exploration

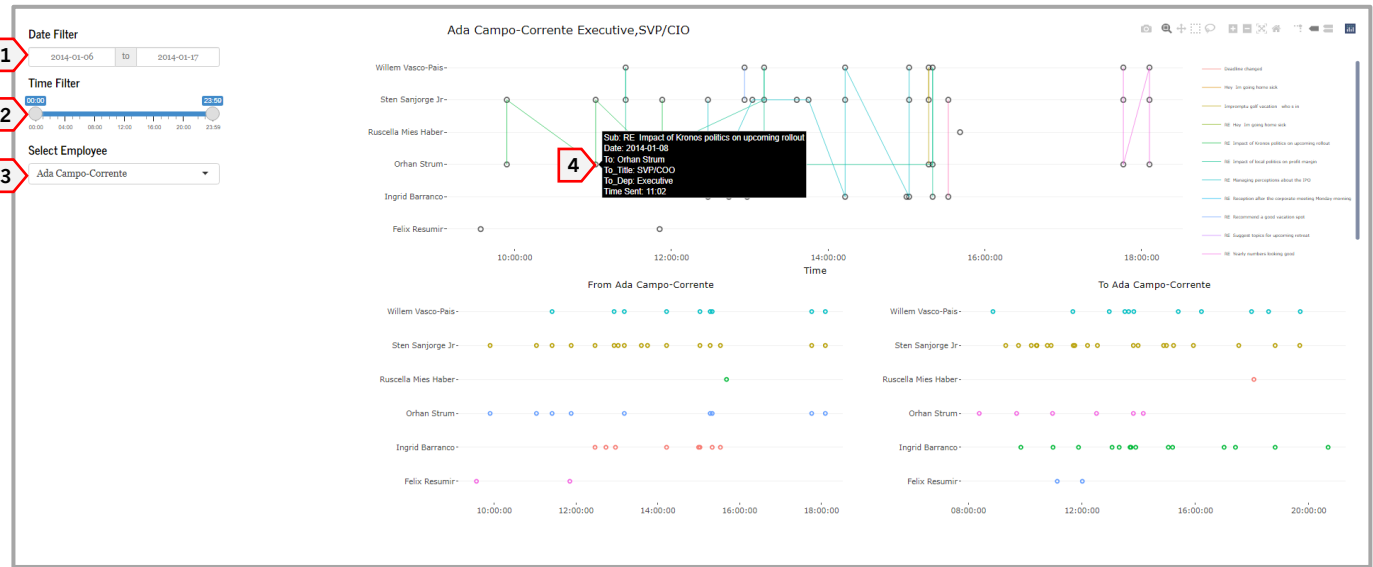
Transactions - For Exploration

Individual Card Transactions

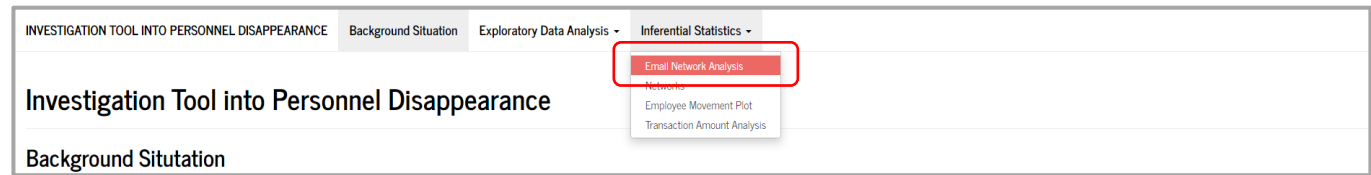
Employees of GasTech

Email Correspondence

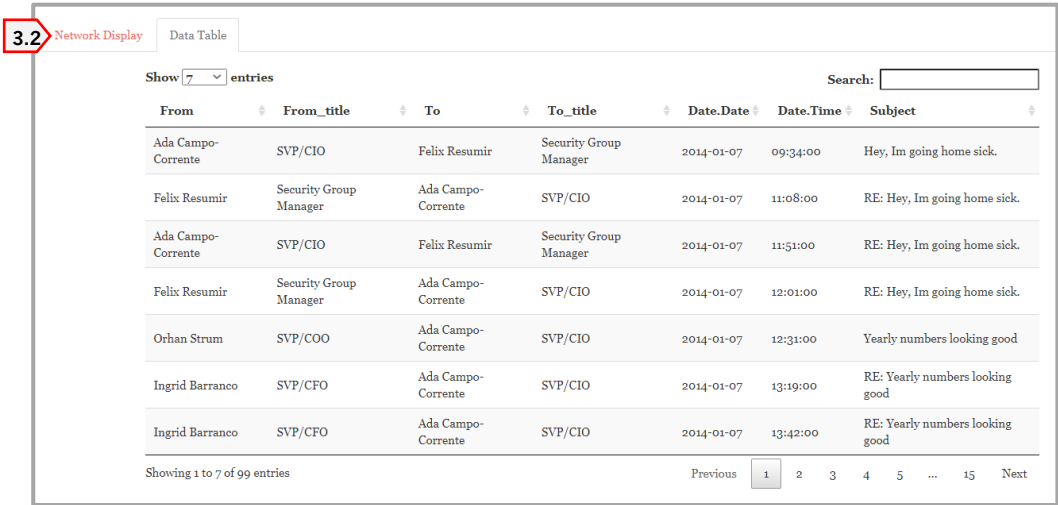
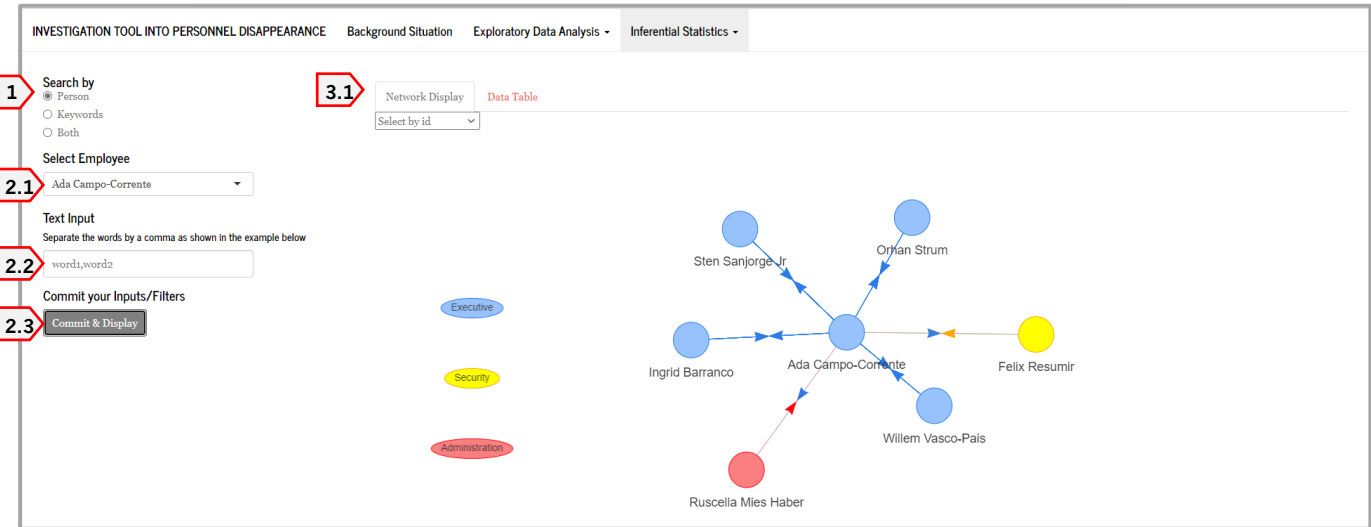
Under the tab **Exploratory Data Analysis**, select **Employees of GasTech**



Email Network Analysis :: Inferential Statistics



Under the tab **Exploratory Data Analysis**, select **Employees of GasTech**



- [1] Select what you wish to search the email correspondence by
- [2.1] If it is by Person, select the Employee you wish to search by
- [2.2] If it is by Keyword, key in the words separated by a comma with no spaces
 Select Both if the search is to be done by both parameters
- [2.3] Once you have confirmed the search, click on Commit & Display
- [3.1] The chart on the tab 3.1 shows the visual representation of the emails that your search brought up.
- [3.2] The data table on the next tab is a break down the visual.
 Each row contains the position of the employee, time and date of the email subject
 Each email chain is separated into individual recipients.

Networks :: Inferential Statistics

INVESTIGATION TOOL INTO PERSONNEL DISAPPEARANCE

Background Situation

Exploratory Data Analysis

Inferential Statistics

Investigation Tool into Personnel Disappearance

Background Situation

Email Network Analysis
Networks
Employee Movement Plot
Transaction Amount Analysis

Under the tab **Exploratory Data Analysis**, select **Employees of GasTech**

INVESTIGATION TOOL INTO PERSONNEL DISAPPEARANCE

Background Situation

Exploratory Data Analysis

Inferential Statistics

1 Select the Network to Build By:
* Email Correspondence
o Email Subjects

2 Node Sizing
* None
o Betweenness
o Degree
o Out-Degree
o In-Degree
o Closeness

3 Community Algorithms
* None
o Department
o Cluster Lovain
o Betweenness
o Label Propagation
o Fast Greedy
o Leading Eigenvector

4 Direction of Edges
* Yes, Display

5 Scale the width of Edges
Input minimum edge weight
Min Value = 0.1, Max Value = 5
0.5
Input maximum edge weight
Min Value = 5, Max Value = 15
7

6 Select Layout Algorithm
Kamada-Kawai

7 Highlight Nearest Nodes
o Yes, highlight

8 Color Sub-Graph Nodes by Department
o Yes, color by department

9 Enable Multi-selection of Nodes
o Yes, enable multi-select

Main Network

Click a node on the Main Network below, to see their sub-graph on the right

Select by group

1.1

1.2

Sub Network

[1] Select the network that you want to display. The first option, Email Correspondence plots the email links between the employees of the GasTech. The second option, plots the text network.

[1.1] The subgraph is generated on click. Click a node you wish to explore, and it will plot the subgraph on click.

[1.2] The graphs can be re-sized using the mouse scroll.

[2] The node can be size according the various algorithms show. The example below shows the nodes sized by the betweenness algorithm. We can see that Felix Resumir has the highest betweenness in the network.

Main Network

Click a node on the Main Network below, to see their sub-graph on the right

Select by group

Executive

Facilities

Engineering

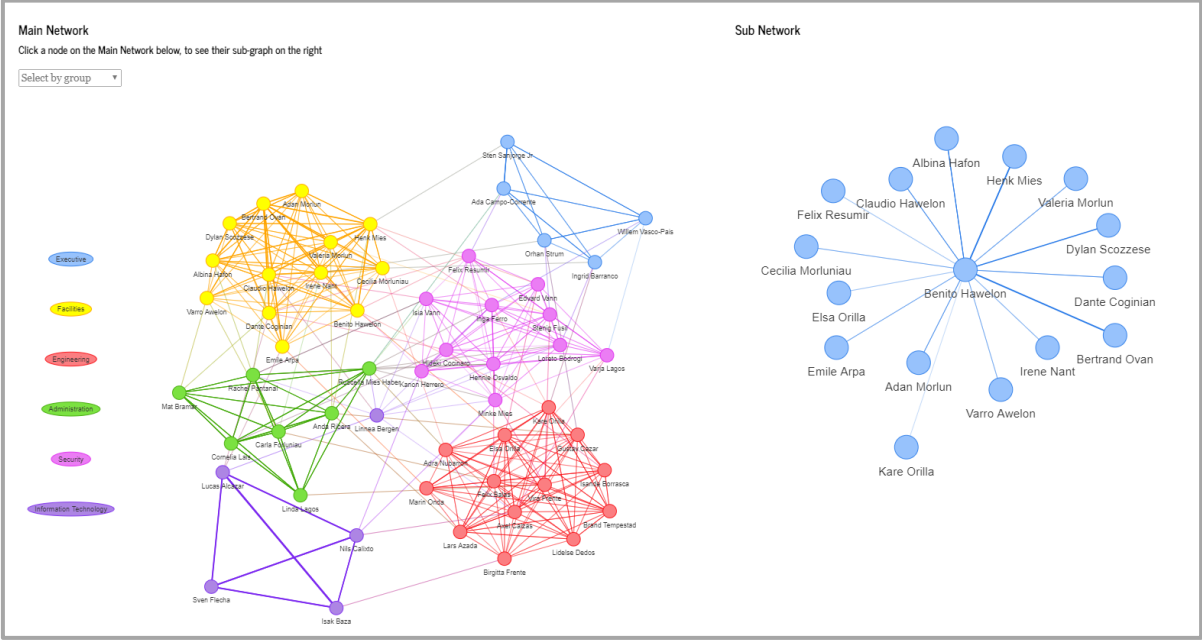
Administration

Security

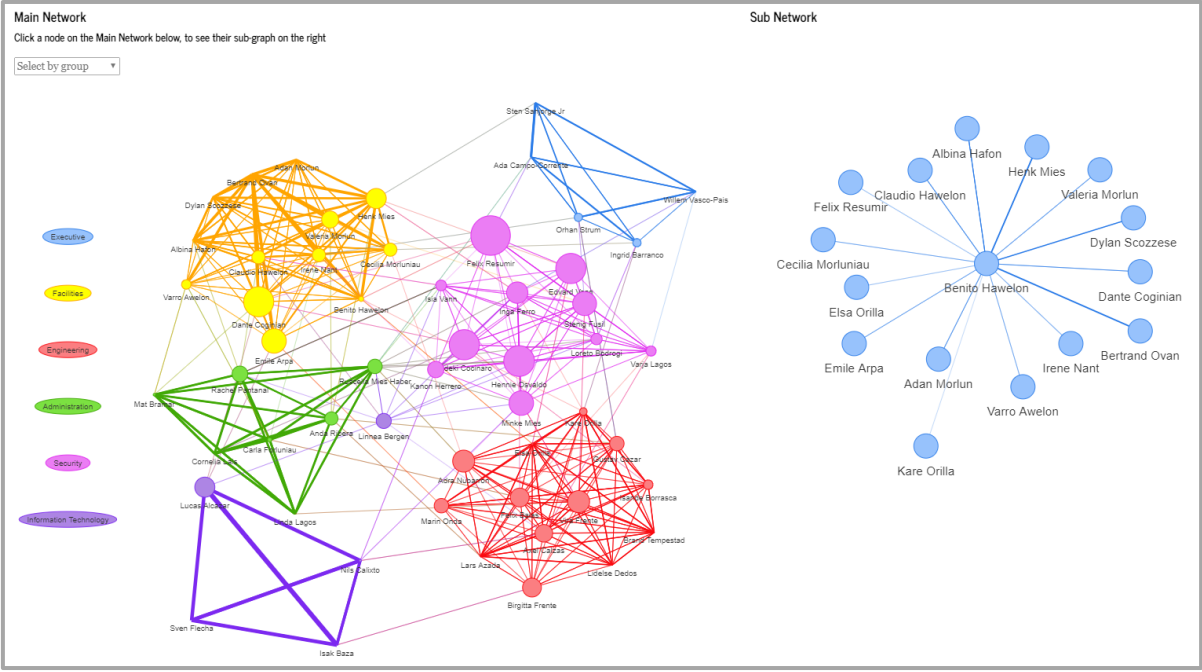
Information Technology

Sub Network

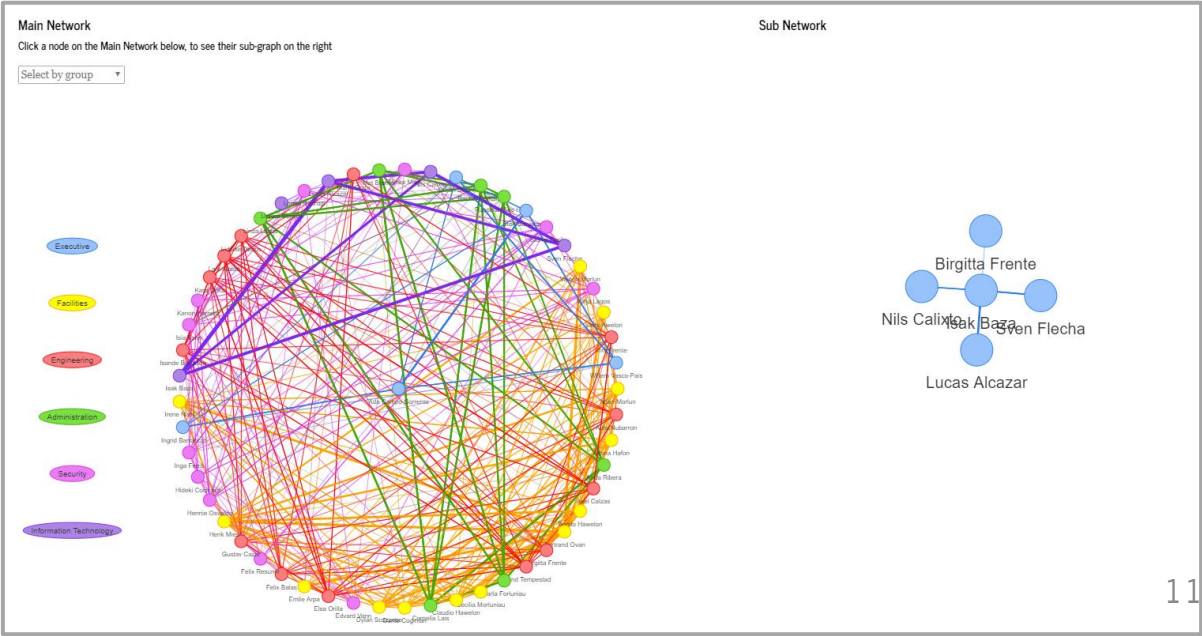
- [3] The communities can be mapped by the given algorithm. Here “None” refers to the department, i.e. the nodes will be coloured by their membership in the department .
- [4] Select if the direction of the edges would like to be shown
The screenshot below shows the graph with the direction de-selected. Note that the node clicked was that of Benito Hawelon.



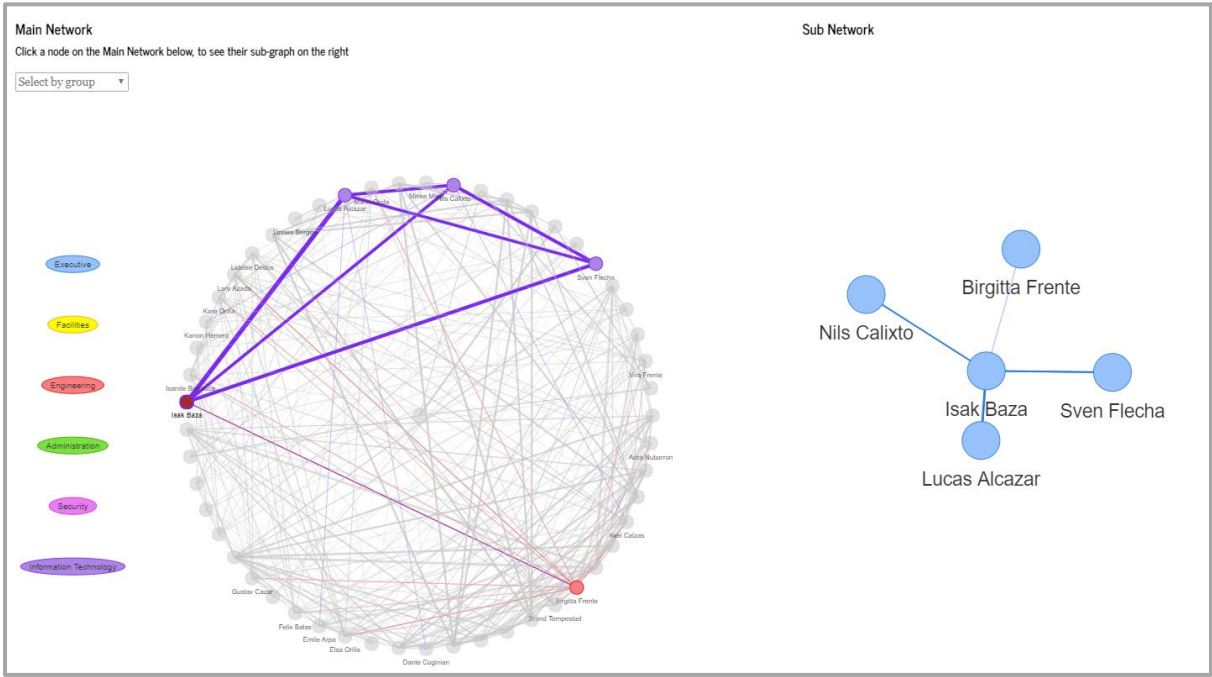
- [5] The edges have been sized by weights i.e. the no of conversations between the employees. The default settings is set to low, but it can be increase to see employees that frequently communicate



- [6] Select the Layout Algorithm. The following shows the graph with a star layout.

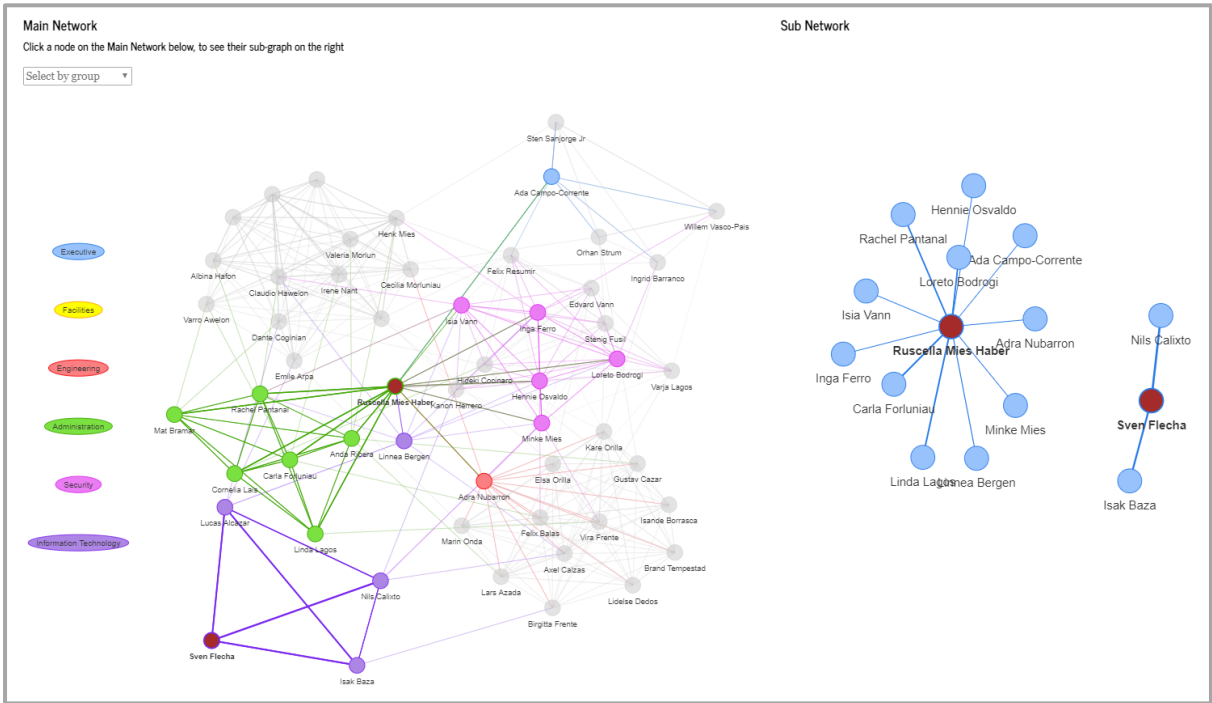


[7] Checking this box will highlight the nodes direct neighbours. The example below shows how a relatively messy layout can be made useful.

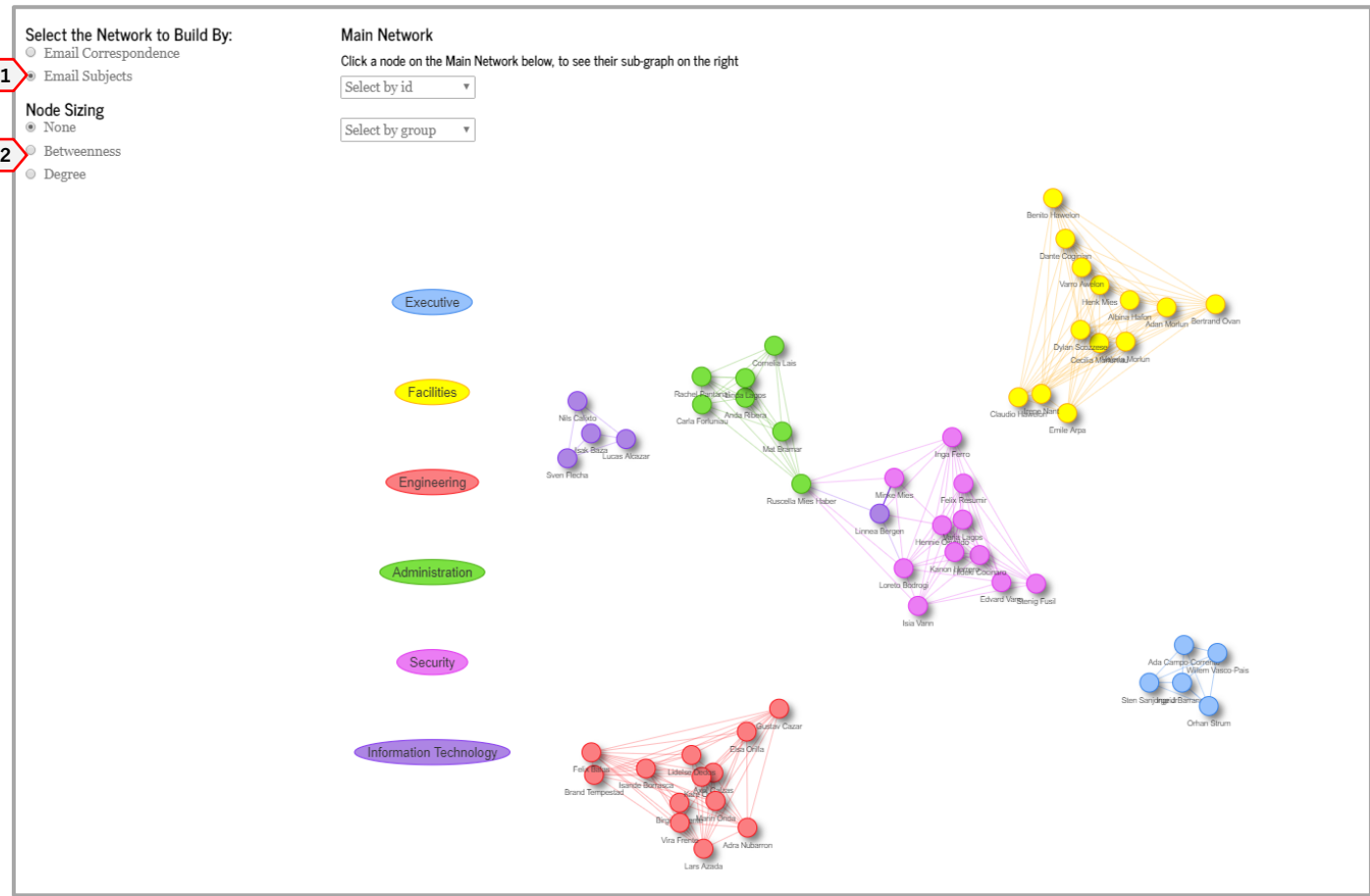


[8] The subgraphs are not colored, this option will help to color it by the department.

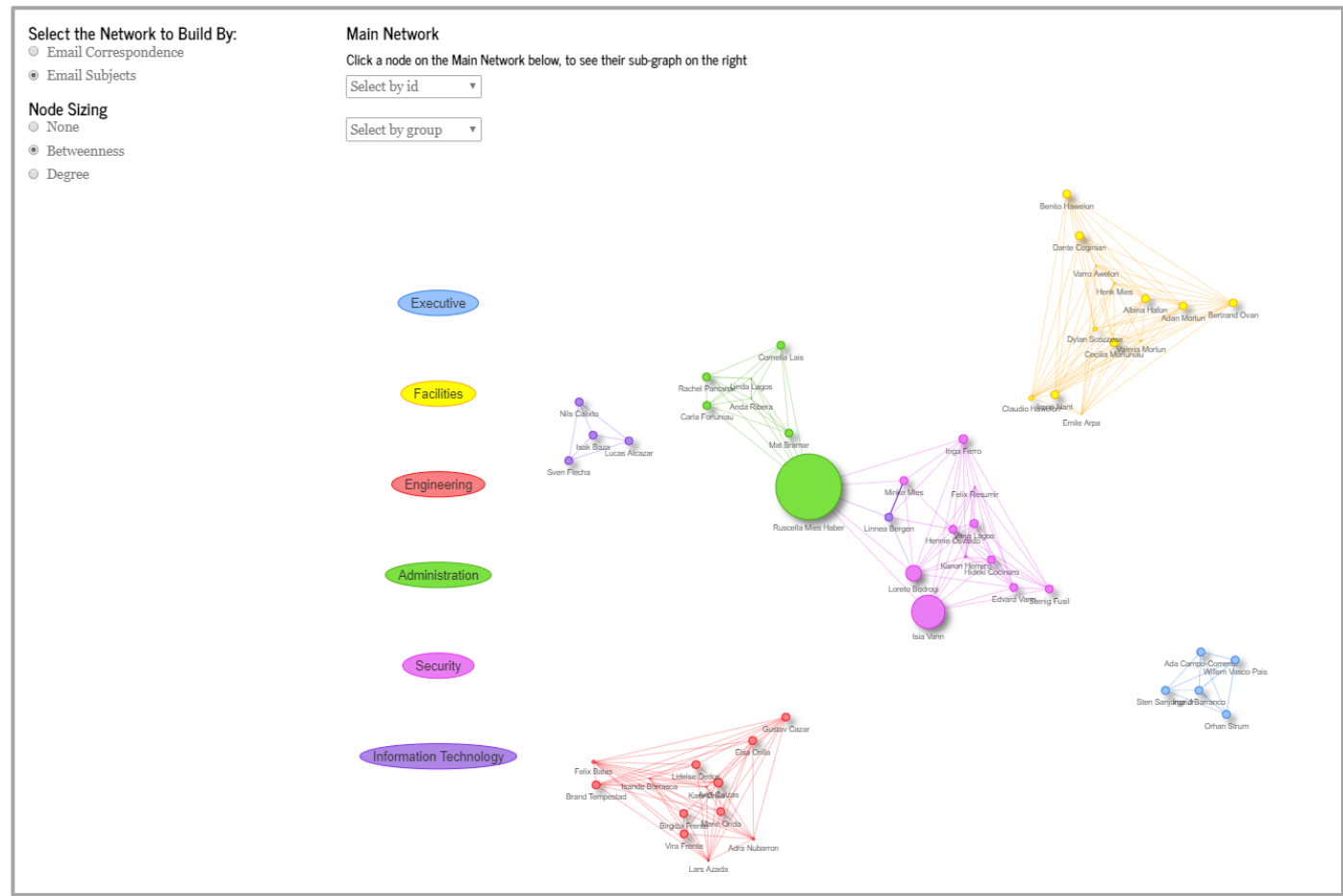
[9] With multiselect enabled, the user can select multiple nodes with a CTRL+click. The subgraphs will also show the central nodes and their connections selected.



Networks :: Inferential Statistics



[1] Email Subjects will give the text network. This option will take time to load.



[2] As previously described, the nodes can be resized by 2 algorithms, degree and betweenness.

Employee Movement Plot :: Inferential Statistics

INVESTIGATION TOOL INTO PERSONNEL DISAPPEARANCE

Background Situation

Exploratory Data Analysis

Inferential Statistics

Email Network Analysis

Networks

Employee Movement Plot

Investigation Tool into Personnel Disappearance

Background Situation

Under the tab **Inferential Statistics**, select **Employee Movement Plot**

Personnel Movement Plot

Employee Name

Hideki Cocinaro

Date

2014-01-06

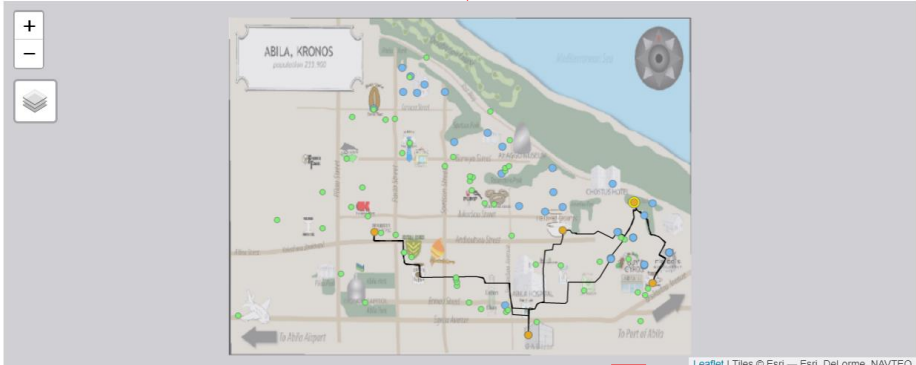
Parking Duration (in minutes)

3

5

15

4



Locations Visited

5

Showing 10 entries

Search:

Order	Arrive	Depart	Duration	Lat	Long	Location
1	2014-01-06T07:11:16Z	2014-01-06T07:44:01Z	32M 45S	36.0637	24.8859	
2	2014-01-06T07:51:09Z	2014-01-06T12:28:01Z	4H 36M 52S	36.048	24.8796	GASTech
3	2014-01-06T12:36:17Z	2014-01-06T14:12:01Z	1H 35M 44S	36.0634	24.851	Robert and Sons
4	2014-01-06T14:23:01Z	2014-01-06T17:12:01Z	2H 49M 0S	36.048	24.8796	GASTech
5	2014-01-06T17:18:41Z	2014-01-06T18:52:01Z	1H 33M 20S	36.0678	24.8991	Home
6	2014-01-06T18:55:19Z	2014-01-06T20:37:01Z	1H 41M 42S	36.0558	24.9026	

Showing 1 to 6 of 6 entries

Previous

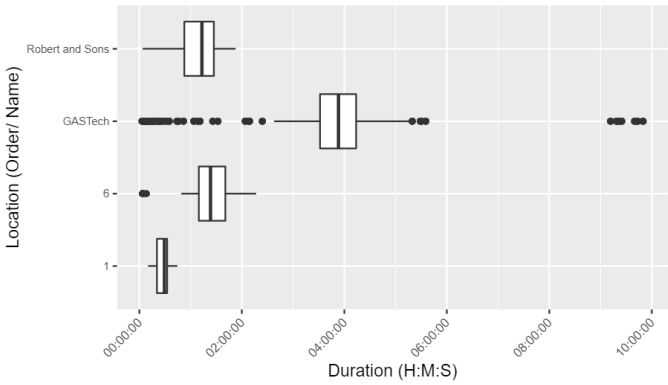
1

Next

Visitation Data

6

Time Spent at Location



- [1] Select the GASTech Employee.
- [2] Select the specific date to explore.
- [3] Provide the definition for parking duration (3 – 15 mins) This would determine the number of locations he/she had been too. If 5 mins were to be selected, that would mean that any time difference of 5 mins or more would be treated as a stop..
- [4] The map, with the corresponding routes would appear. Sky Blue would represent the homes of GASTech Employees, Light Green to represent locations, while the orange dot would indicate the locations that the employee had been to. The dot with a yellow backing indicates the residential home of the employee.
- [5] The order of the locations would be reflected on the datatable, time of arrival, departure time and the duration he/she had spent at the location.
- [6] Based on the locations listed, it would indicate the duration statistics for the locations, in the form of a box plot, to allow the investigator to infer if the stay at the location was longer than usual.

Transaction Amount Analysis :: Inferential Statistics

INVESTIGATION TOOL INTO PERSONNEL DISAPPEARANCE

Background Situation

Exploratory Data Analysis

Inferential Statistics

- Email Network Analysis
- Networks
- Employee Movement Plot
- Transaction Amount Analysis

Investigation Tool into Personnel Disappearance

Background Situation

Under the tab **Inferential Statistics**, select **Transaction Amount Analysis**



- [1] Select the category of locations you would like to display boxplots and compare the transaction amounts for. The locations are categorized so that the transactions of the same type are compared. For e.g. we can observe that for the eateries, higher amounts are spent at Abila Zacharo compared to Coffee Cameleon. Note that the x-axis is a log-scale.
- [2] You may also hover along the box plots to view the distribution of each individual transaction, appearing as a tooltip. Outliers are also highlighted in red.
- [3] Select the card type you want to visualize the transaction spread for, and then the card number
- [4] You may hover over each diamond to view the details of the transaction. Each diamond is filled according to the magnitude of the transaction amount. Similar to the boxplot above, the x-axis is also in log-scale.