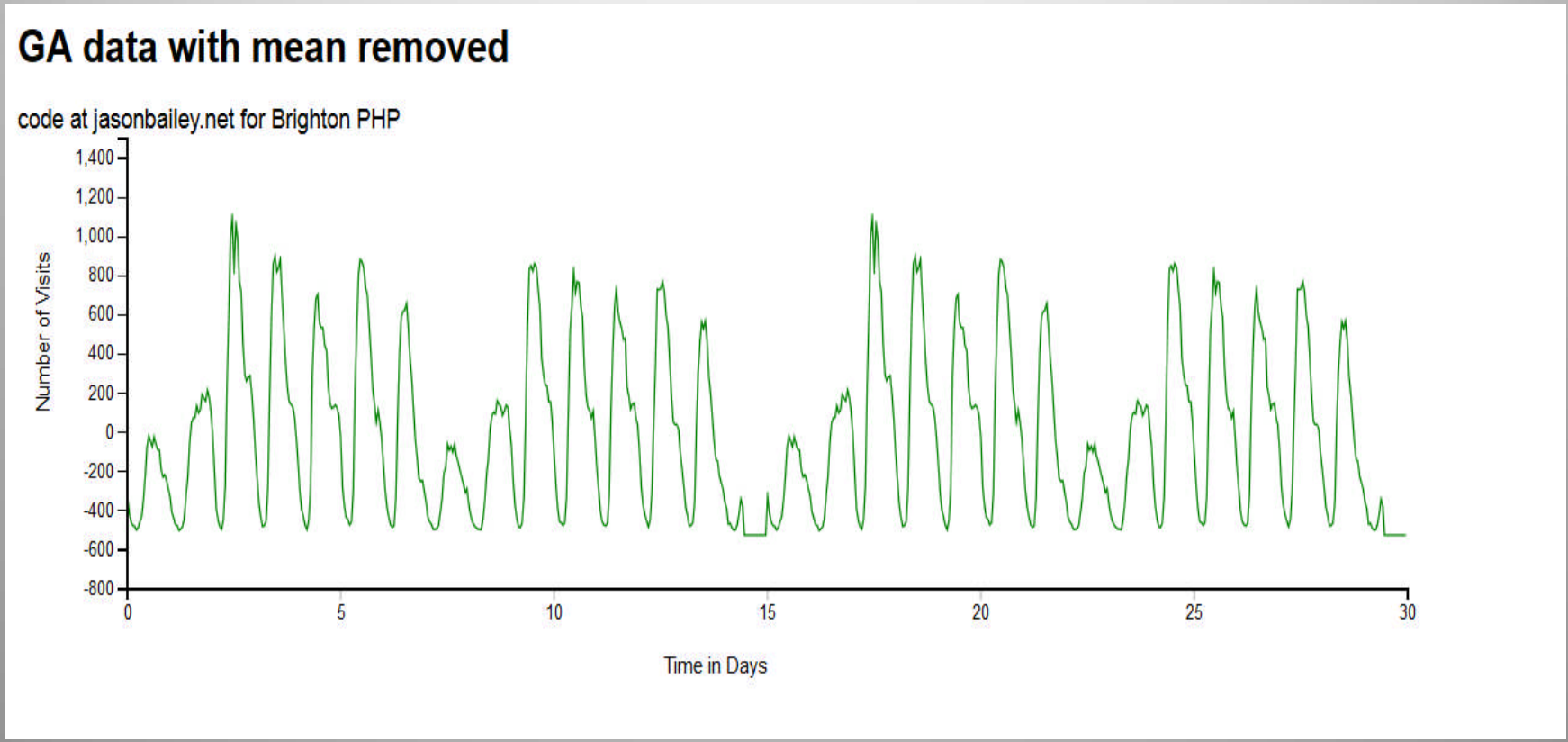
Finding patterns in Google Analytics data using PHP



Jason Bailey

Brighton PHP October 2013

To cover

- Look at Time Series Data
- See data in Time domain (time series) and Frequency domain (using Fourier Transform)
- FFT - Fast Fourier Transform –Examples in Matlab
- PHP to get Data from Google Analytics
- PHP to process the data
- Use d3.js to present the JSON data (WIP)
- WIP - not quite finished (sorry)
- Code at:

http://www.jasonbailey.net/stuff/brighton-php-october-2013-talk/

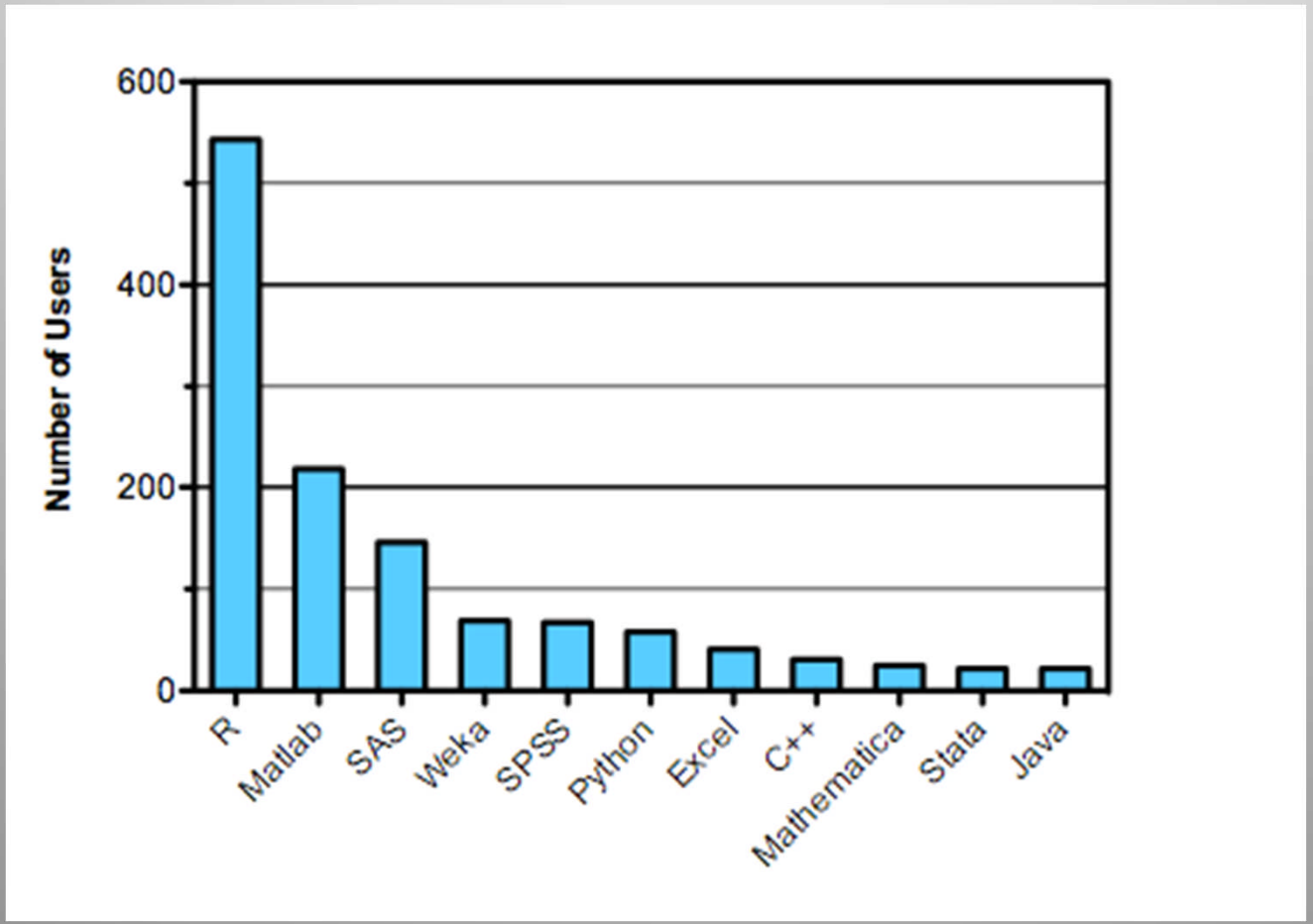
Why look for patterns?

- Prediction of load on server/application based on a model
- When is a good time to introduce a product or downtime
- Model (predict) profit based on demographics

Is this a good way to do it?

- PHP + d3.js – I started thinking no, but ….
- Automated
- During development find good tools to quickly see the results
  - Matlab –expensive
  - R – Free
- Charts/images are out of date quickly –automate to keep data fresh
- Open up scientific data – usually not available online

People on Kaggle use



Source: http://www.kaggle.com/wiki/Software

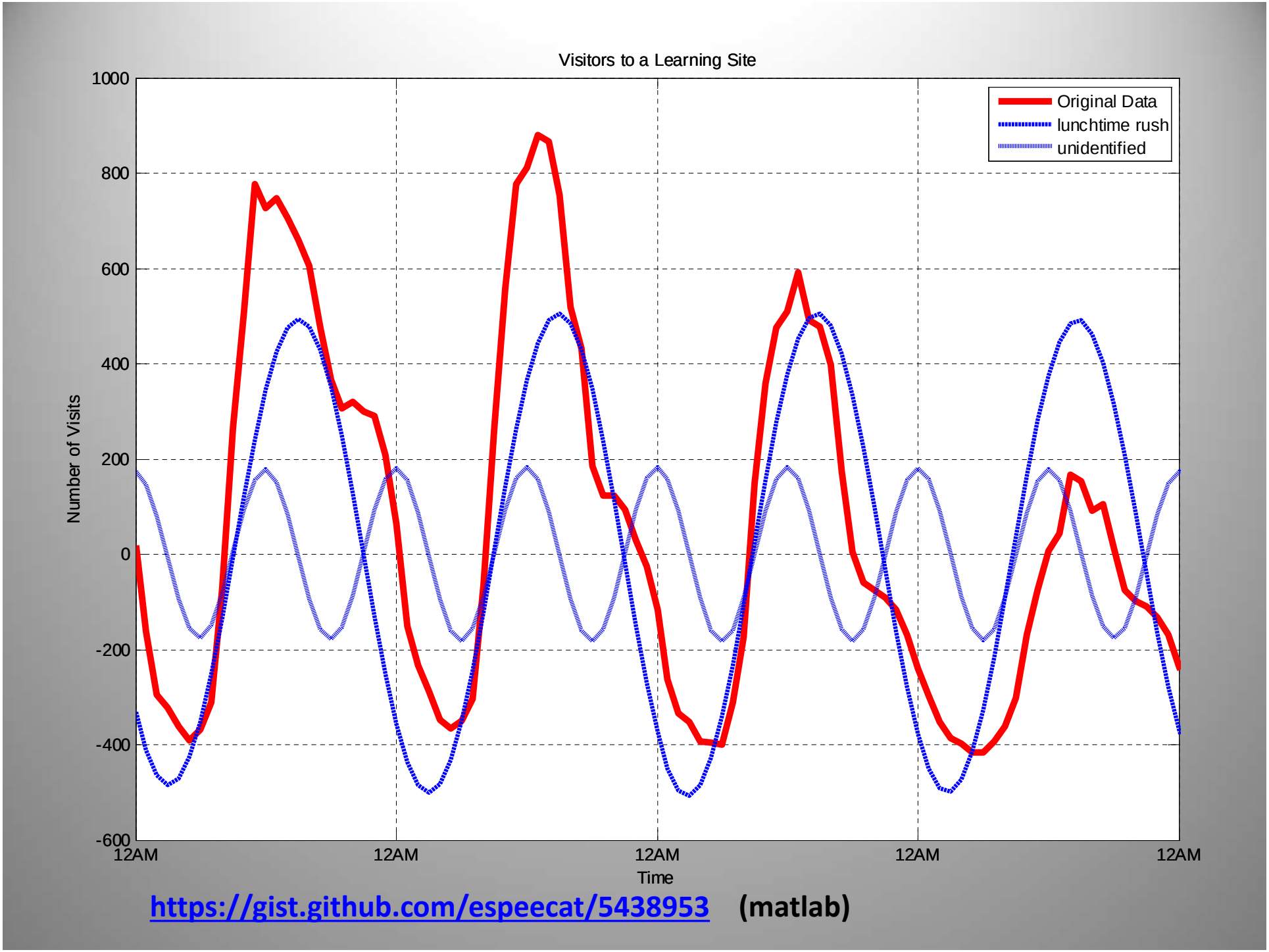
What is Time Series Data

- A sequence of data points
- Typically at successive points in time spaced at uniform time intervals
- Used:

statistics, signal processing, pattern recognition, finance, weather forecasting, earthquake prediction, control engineering and communications engineering

http://en.wikipedia.org/wiki/Time_series

What if we want to extract
a pattern
from time series data?



https://gist.github.com/espeecat/5438953 (matlab)

A sine wave or sinusoid

$$y(t) = A \sin(2\pi f t + \phi)$$

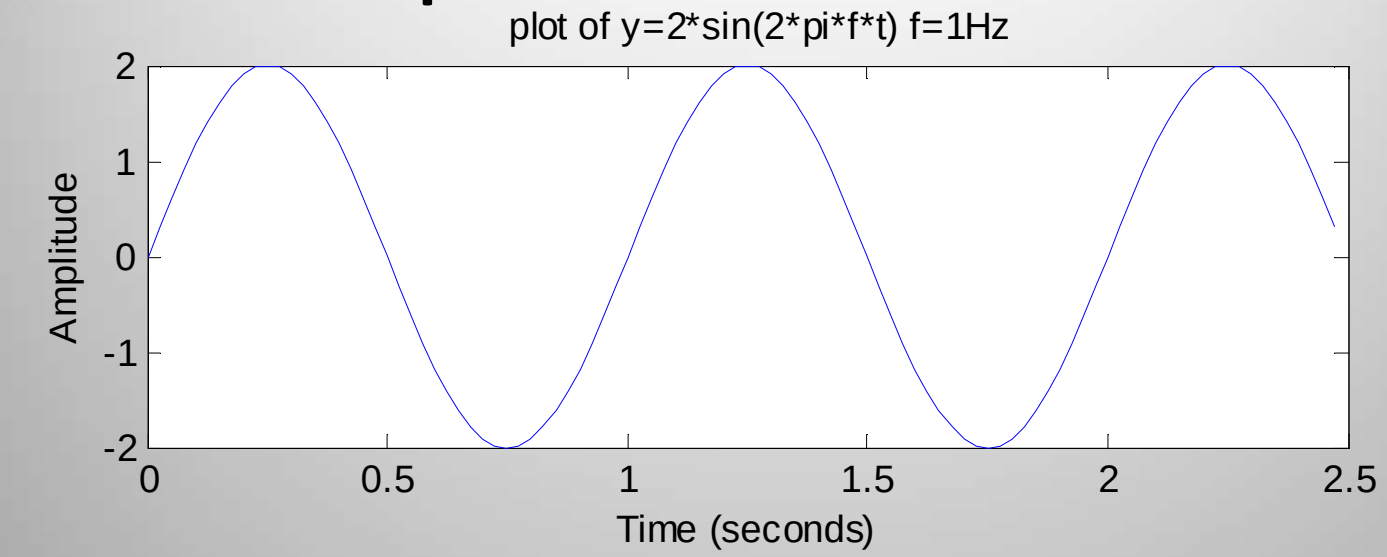
Sometimes $2\pi f$ written as ω

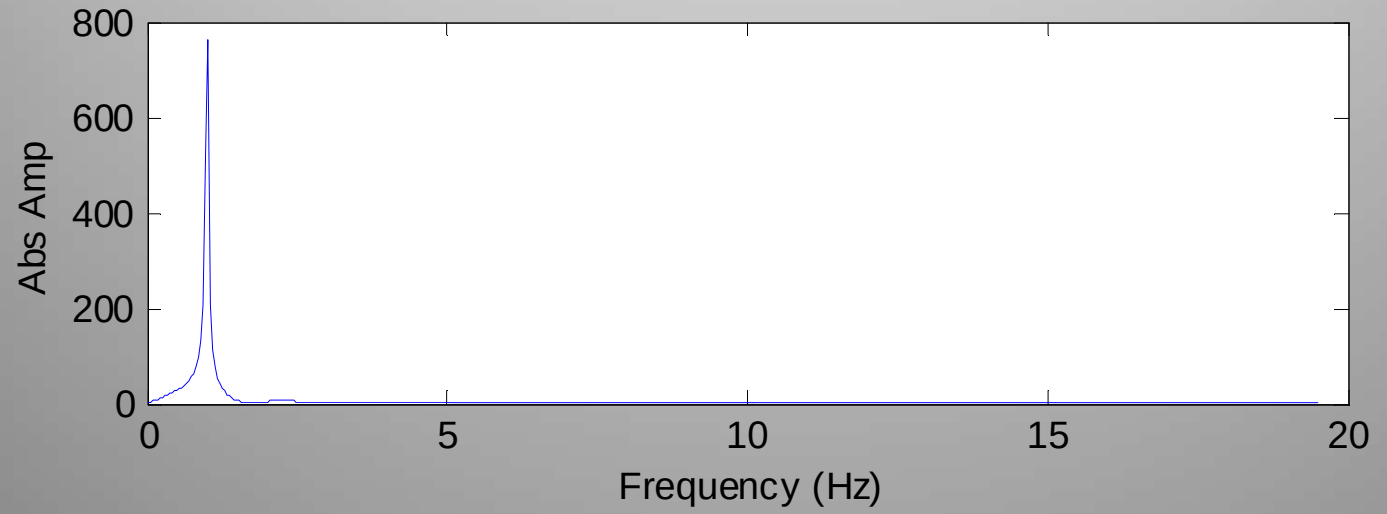
Cosine too

Much better to see it in a graph

- Use a tool like Matlab/R while developing
- Process data and make it look nice using d3.js
- The Fourier Transform splits time series data into discreet frequency components (sinusoids)

An example of a sinusoid and FFT





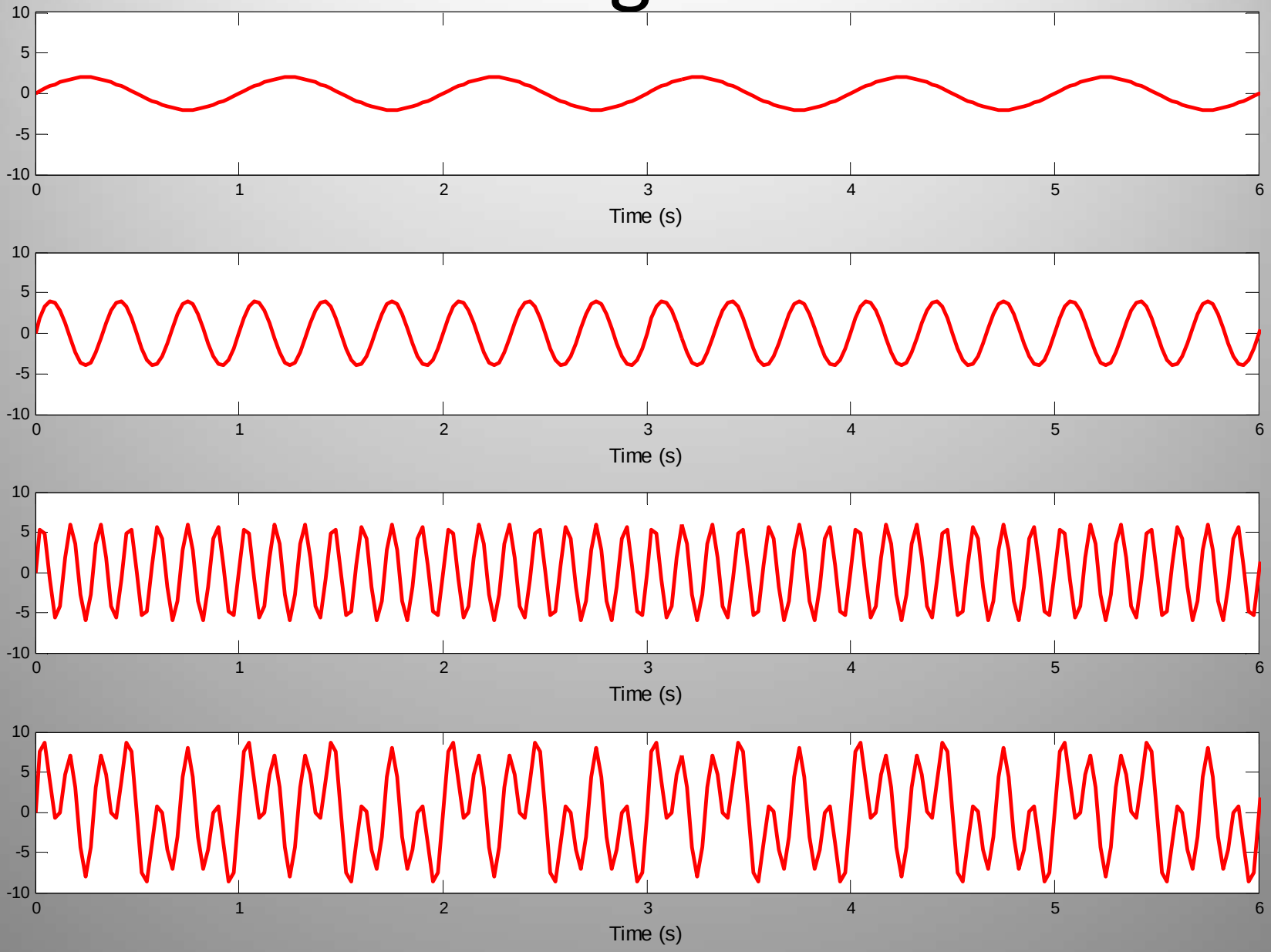
https://gist.github.com/espeecat/5439069 (matlab)

The Fourier Transform (FFT)

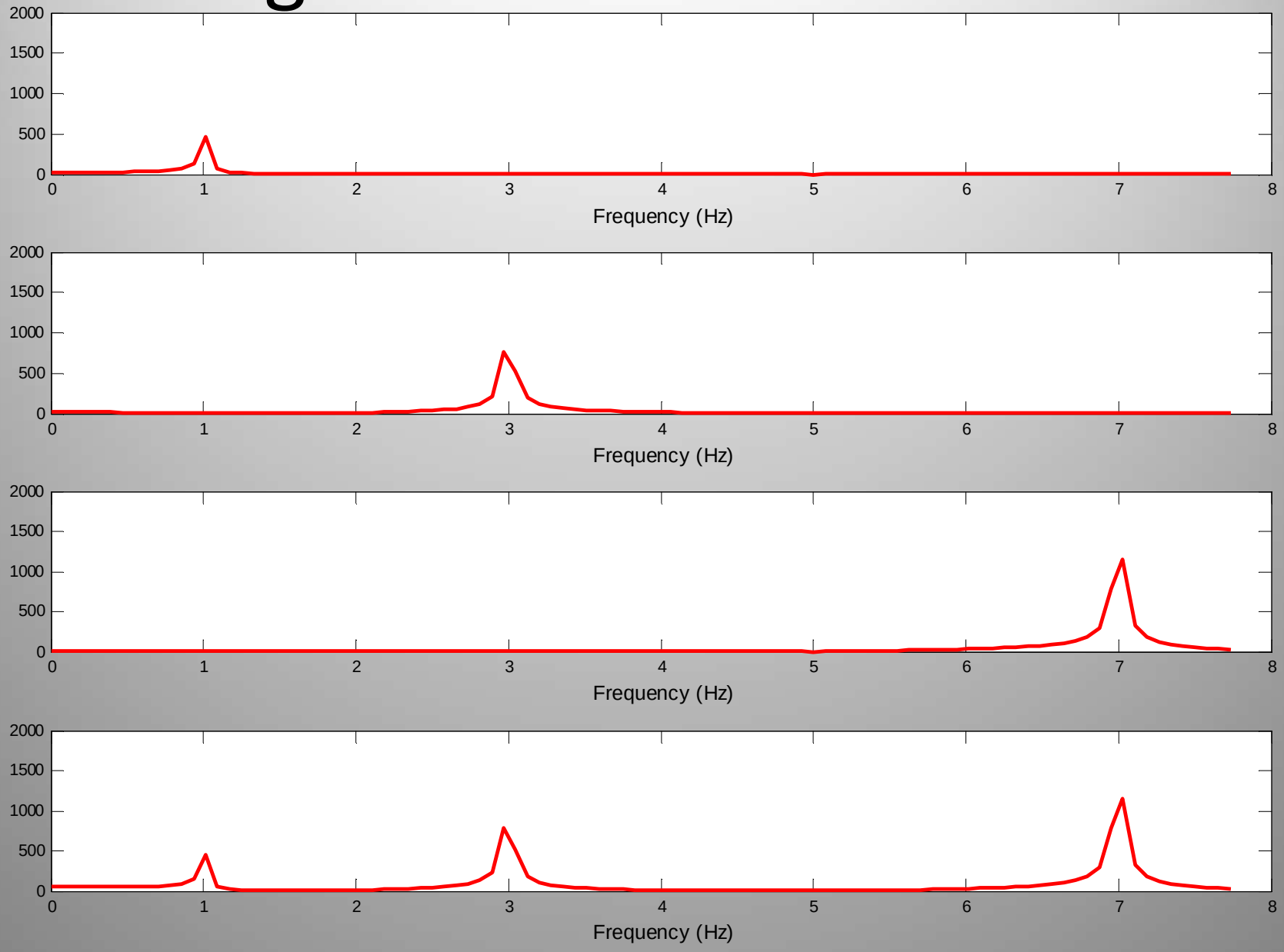
- Based on Fourier Series - represent periodic time series data as a sum of sinusoidal components (sine and cosine)
- (Fast) Fourier Transform [FFT] – represent time series in the frequency domain (frequency and power)
- The Inverse (Fast) Fourier Transform [IFFT] is the reverse of the FFT
- Like graphic equaliser on music player

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nx) + b_n \sin(nx)$$

Combining Sinusoids



Looking at the Fourier Transforms



Applications of Fourier Transform

- Shazam – “finger printing” using Fourier Transforms
- Images – The Gabor Transform for facial recognition
- Filtering data/ Extracting patterns
- Sound processing – discarding sound
- System Identification
  - Transfer function $H(f) = Y(f) / X(f)$

The (Fast) Fourier Transform

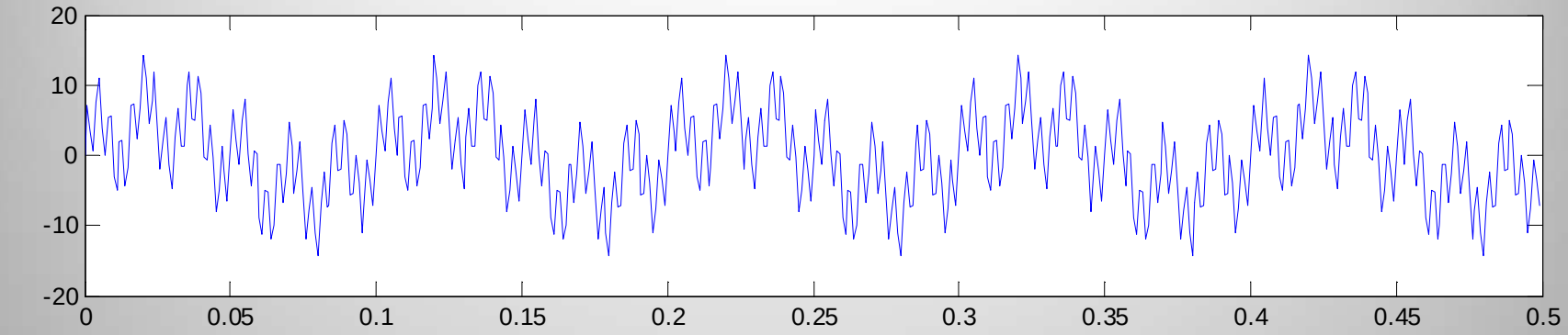
- Discrete-time Fourier Transform –assumes sampled data and limited length
- Implementations available in lots of programming languages e.g. http://www.fftw.org/
- PHP implementation at:

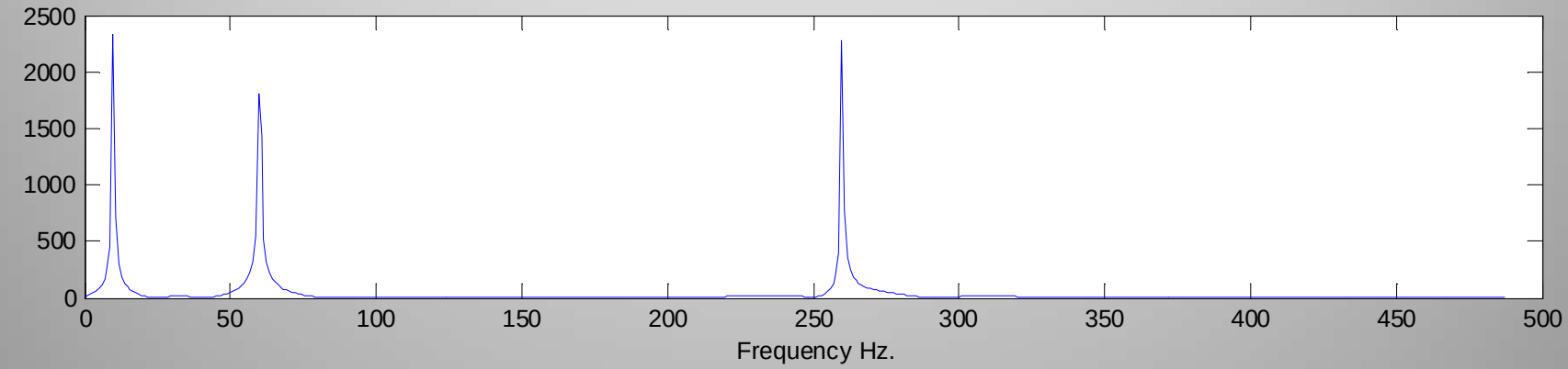
http://www.phpclasses.org/package/6193-PHP-Compute-the-Fast-Fourier-Transform-of-sampled-data.html

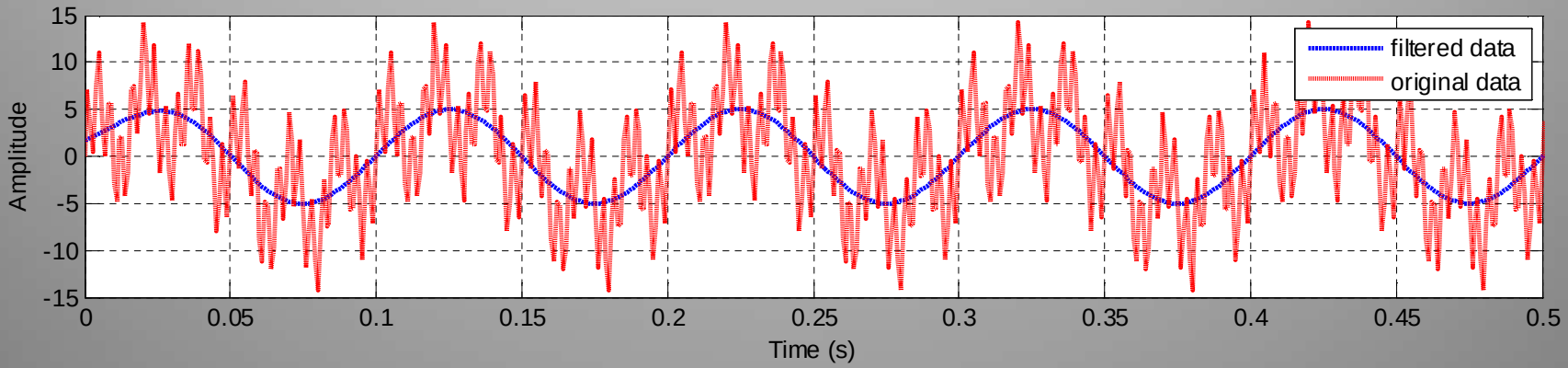
Google Analytics API at:

http://code.google.com/p/google-api-php-client

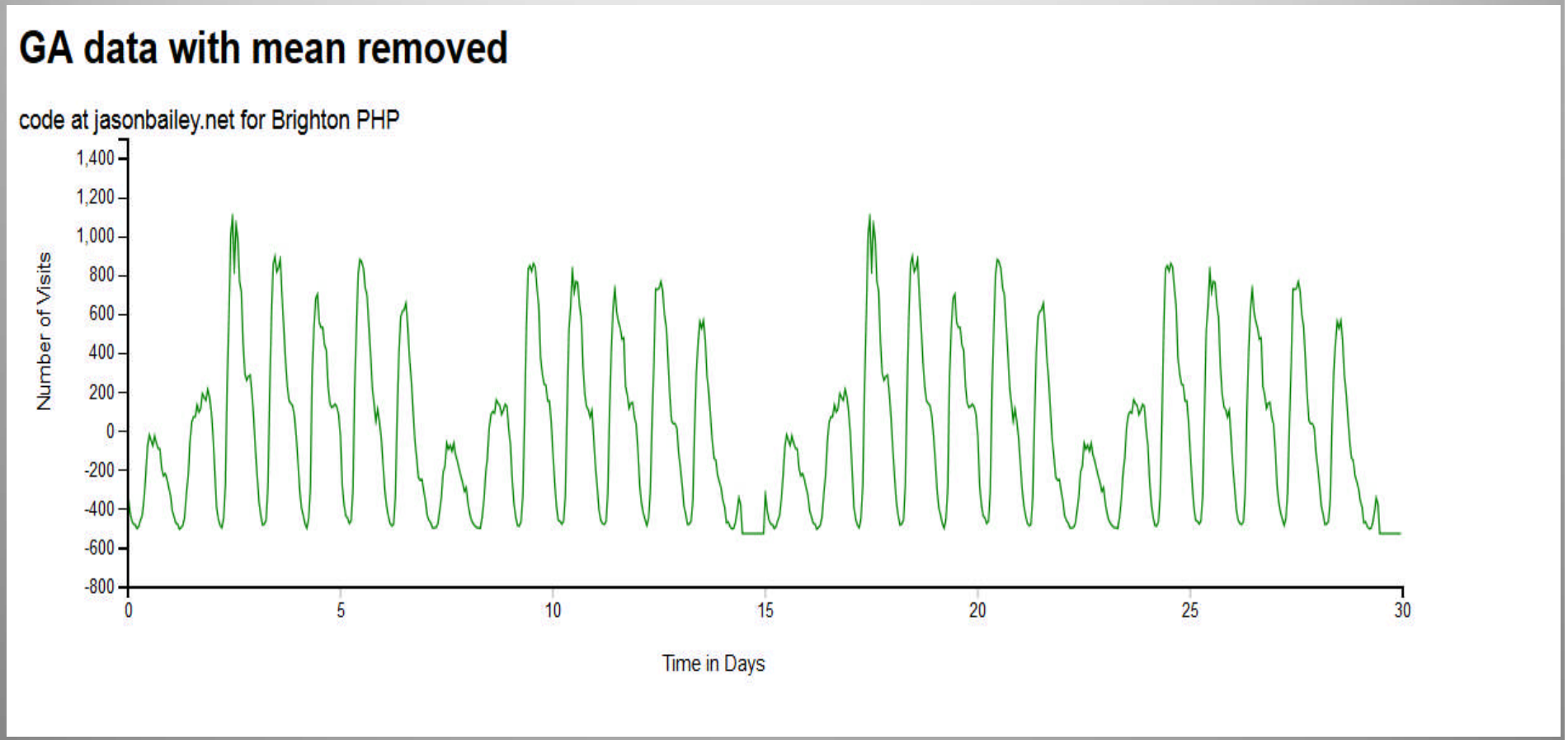
Filtering Time Series Data







Get the Google Analytics data with PHP and save as JSON

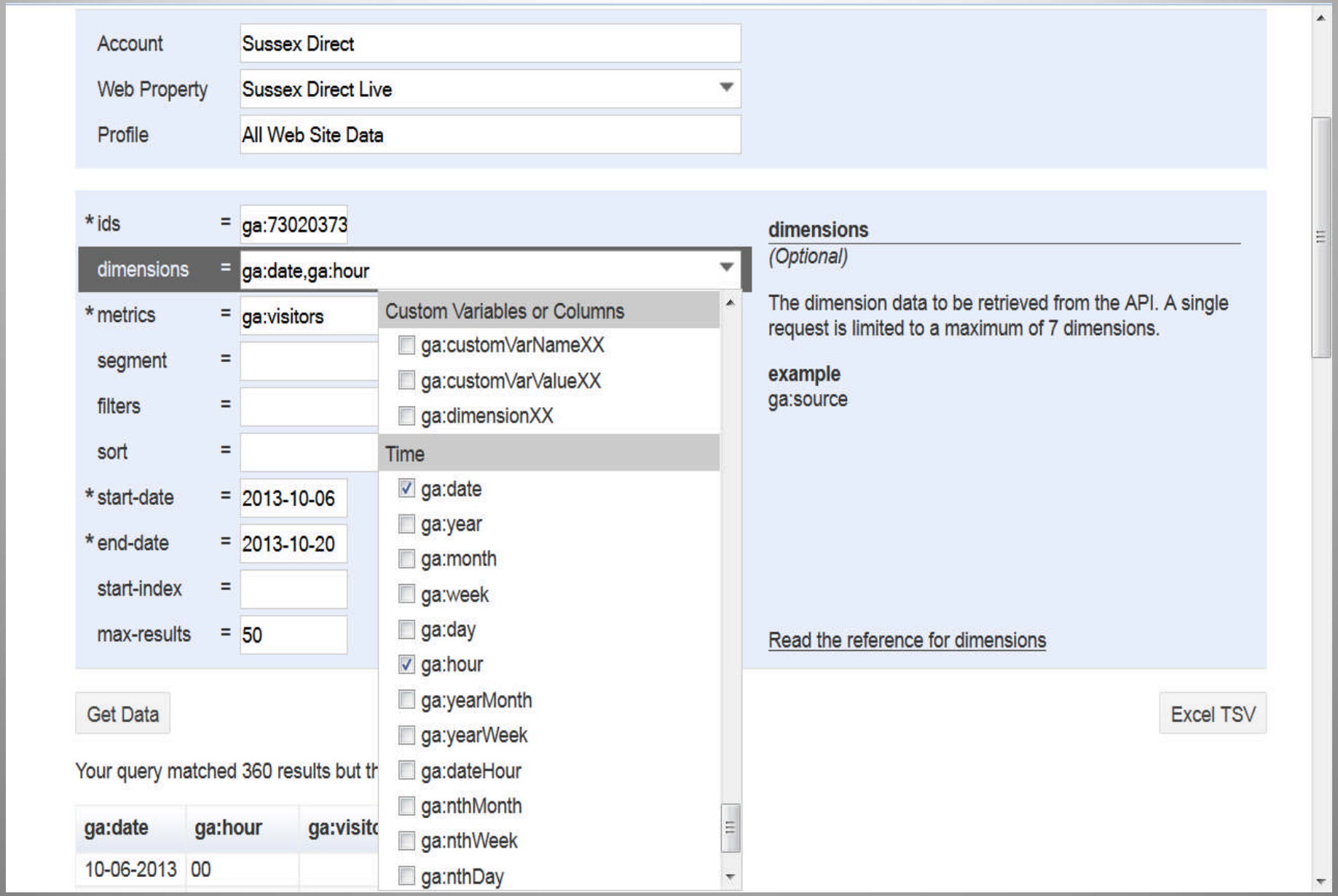


http://erushi.com/brightonphp/timeseries.html

Google Analytics Query Explorer 2

Useful tool

http://ga-dev-tools.appspot.com/explorer/



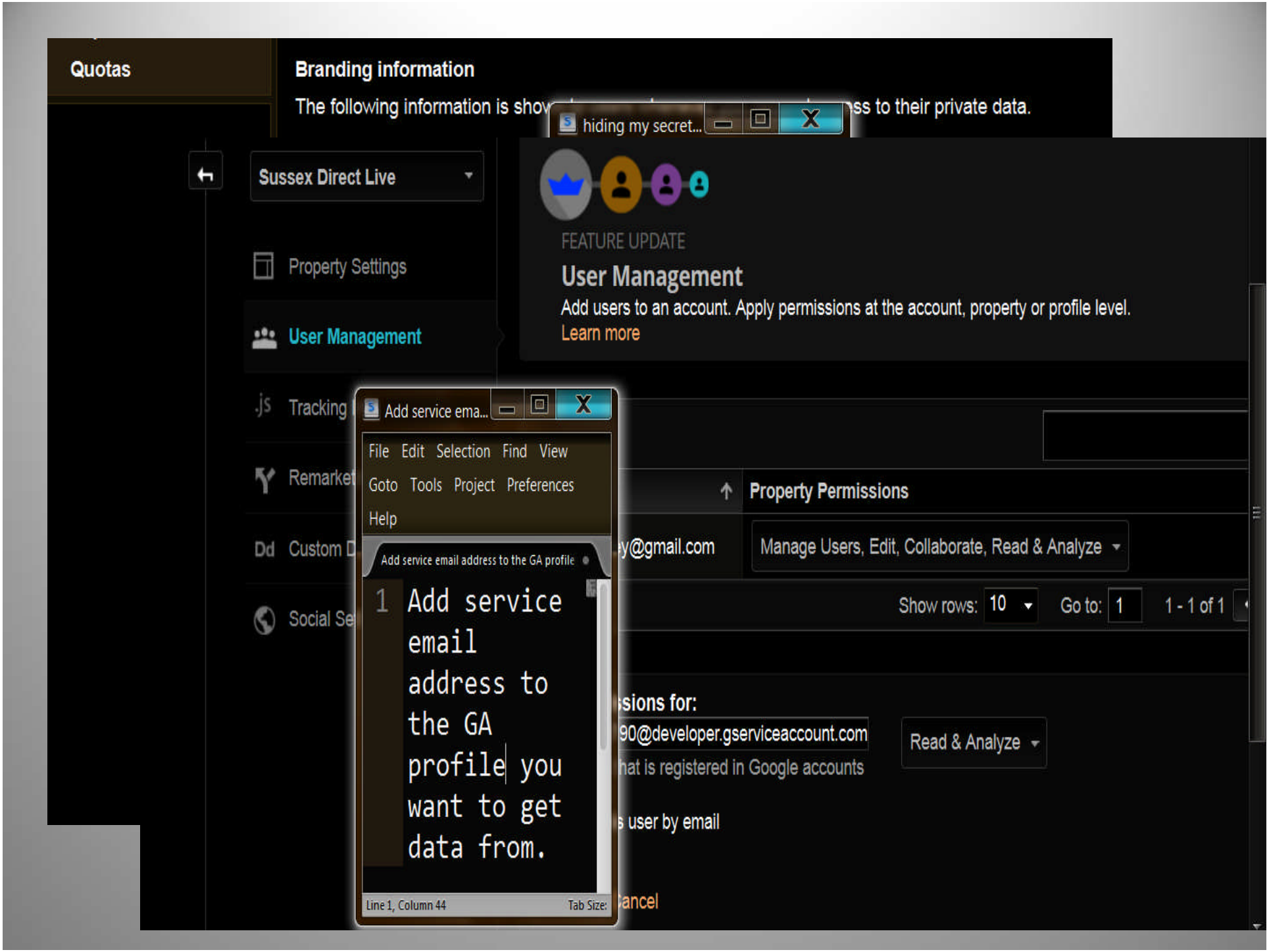
Setup Access to GA for your code

Set up service account at Google API console

Key and Email

Allow service email access to GA

https://code.google.com/apis/console/


Quotas
Branding information
The following information is shov
ss to their private data.
hiding my secret...
Sussex Direct Live
Property Settings
User Management
Tracking
Remarket
Custom D
Social Se
FEATURE UPDATE
User Management
Add users to an account. Apply permissions at the account, property or profile level.
Learn more
Add service ema...
File Edit Selection Find View
Goto Tools Project Preferences
Help
Add service email address to the GA profile
1 Add service email address to the GA profile you want to get data from.
Line 1, Column 44
Tab Size:
Property Permissions
y@gmail.com
Manage Users, Edit, Collaborate, Read & Analyze
Show rows: 10
Go to: 1
1 - 1 of 1
sions for:
90@developer.gserviceaccount.com
Read & Analyze
hat is registered in Google accounts
s user by email
ancel

Code for illustration

I didn’t go through each line

API at:

http://code.google.com/p/google-api-php-client

```php
// Initialise the Google Client object
$client = new Google_Client();
$client->setApplicationName('Brighton PHP talk');
$client->setAssertionCredentials(
        new Google_AssertionCredentials(
        '3xxxxxxxxx0@developer.gserviceaccount.com', array('https://www.google
        )
);
$client->setClientId('3xxxxxxxx0.apps.googleusercontent.com');
$client->setAccessType('offline_access');
$analytics = new Google_AnalyticsService($client);

$analytics_id = 'ga:73020373';  // http://productforums.google.com/forum/#!top
// get data for the last 2 weeks
$lastWeek = date('Y-m-d', strtotime('-2 week'));
$today = date('Y-m-d');
// Test connection
try {
    $results = $analytics->data_ga->get($analytics_id, $lastWeek, $today, 'ga:
    echo '<b>Number of visits this week:</b> ';
    echo $results['totalsForAllResults']['ga:visits'];
```

Code for illustration

I didn’t go through each line

```
63
64 $metrics = "ga:visits"; // number of visits
65 $dimensions = "ga:date,ga:hour"; // by days, hours
66 $sort = "ga:date,ga:hour"; // order by date, time
67 $optParams = array('dimensions' => $dimensions, 'sort' => $sort);
68 try {
69     $results = $analytics->data_ga->get($analytics_id, $lastWeek, $today, $met
70 } catch (Exception $e) {
71     echo 'There was an error : - ' . $e->getMessage();
72 }
73 $data = $results ['rows']; // This is the time series data we want
```

```
$results =
    $analytics->data_ga
        ->get($analytics_id, $lastWeek, $today, $metrics, $optParams);
```

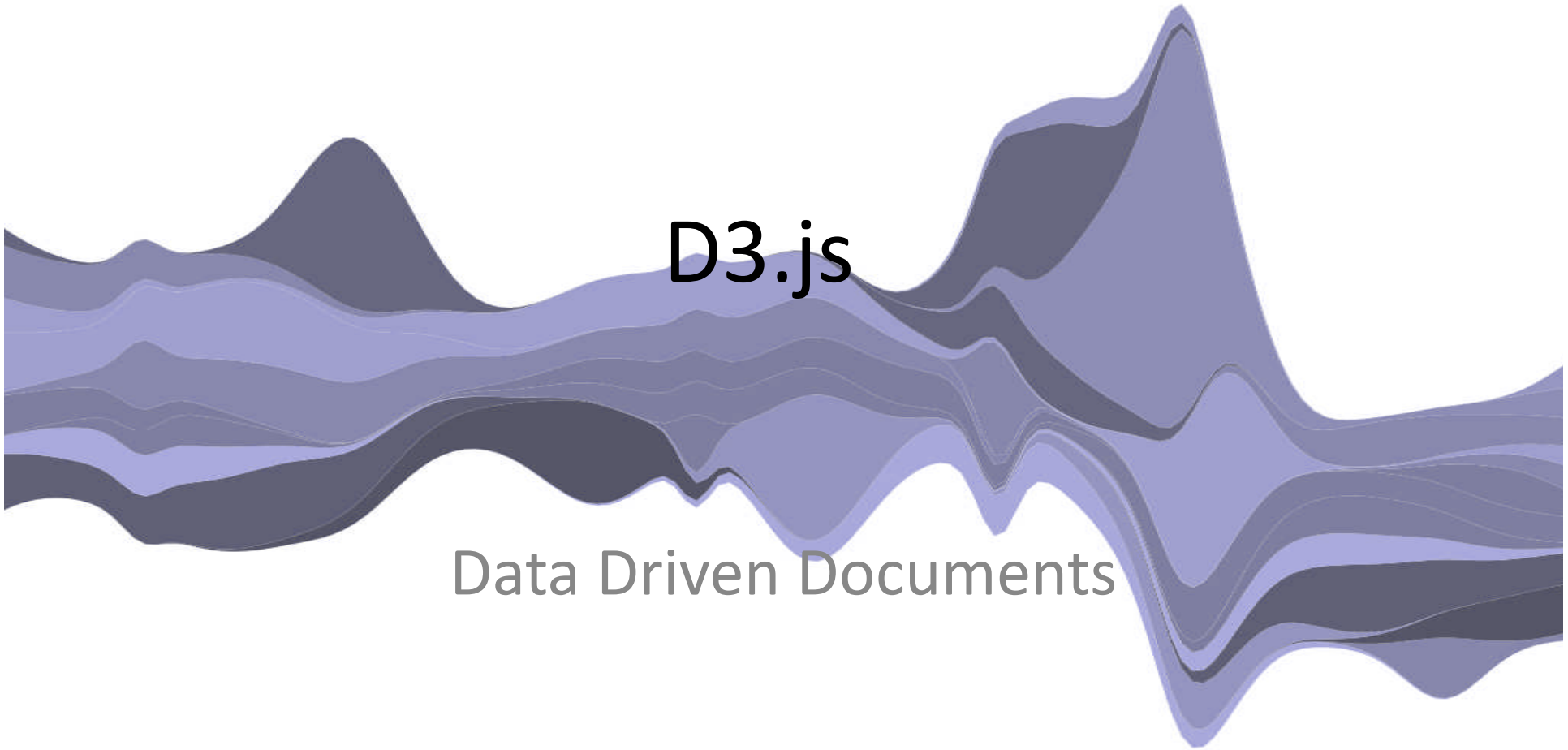
Do the FFT with PHP

Code for illustration

I didn’t go through each line

```php
$nfft = 256;  // n point FFT: Good example in Matlab at:
// http://www.mathworks.co.uk/help/matlab/ref/fft.html
//zero mean GA data
```

```php
$fft = new FFT($nfft);
// create power/abs fft
$data = json_decode($string);
// Calculate the FFT of the function $f
$w = $fft->fft($data);
$power = $fft->getAbsFFT($w);

// save abs value of absfft.json for fft.html
echo 'data for fft.html<br />';
$powerJson = json_encode($power);
```



D3.js

Data Driven Documents

Code for illustration

I didn’t go through each line

```
<script type="text/javascript">
    var w = 500;
    var h = 50;
    var dataset = [ 5, 10, 15, 20, 25 ];
    //Create SVG element
    var svg = d3.select("body").append("svg")
                .attr("width", 500).attr("height", 50);

    var circles = svg.selectAll("circle").data(dataset).enter()
        .append("circle");

    circles.attr("cx", function(d, i) {
                return (i * 50) + 25;
            })
           .attr("cy", h/2).attr("r", function(d) {
                return d;
           });

</script>
```

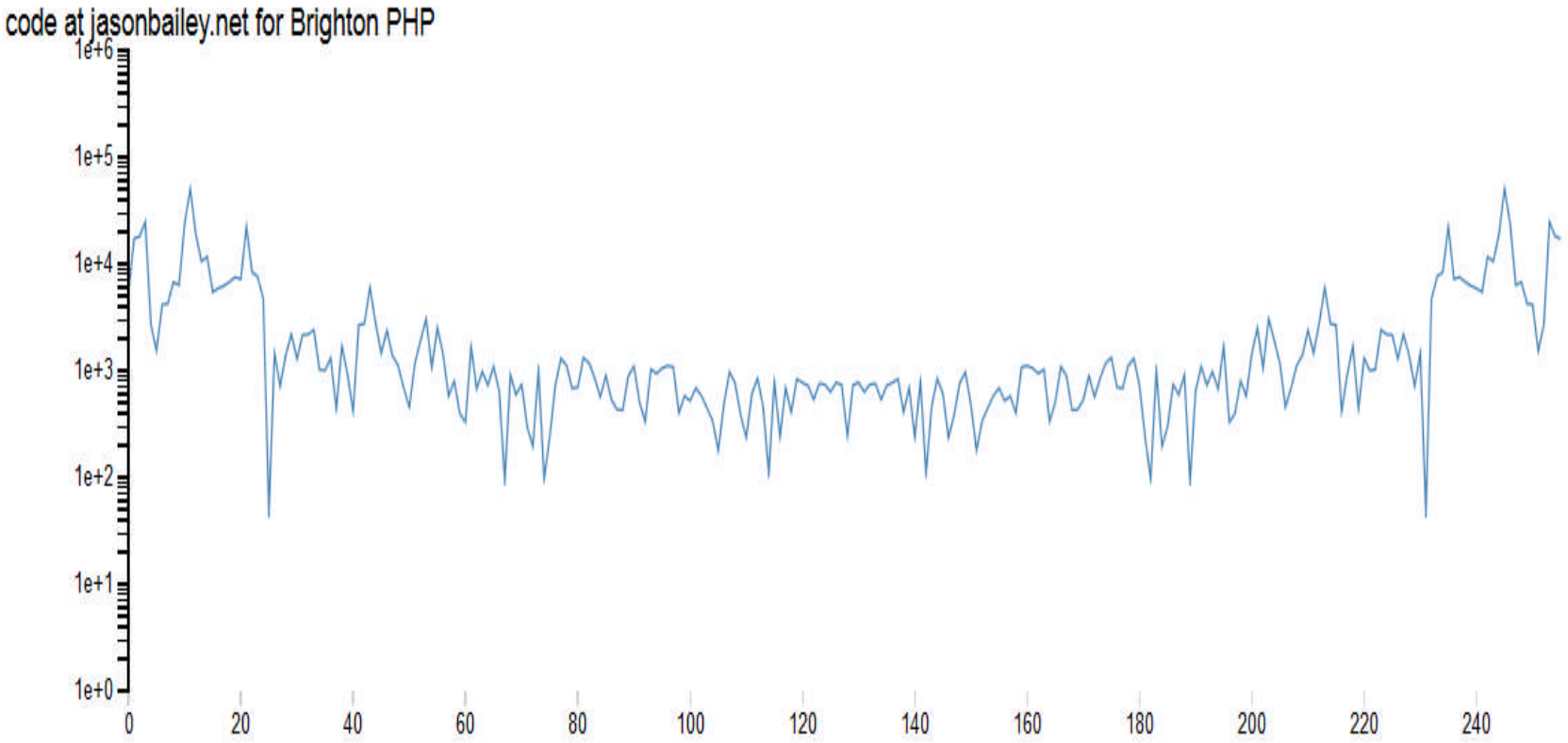
http://knowledgestockpile.blogspot.co.uk/2012/01/understanding-selectall-data-enter.html

http://alignedleft.com/content/03-tutorials/01-d3/110-drawing-svgs/3.html

Present the data with d3.js

The absolute value of the Fast Fourier Transform

Absolute value of FFT components (log)



http://erushi.com/brightonphp/fft.html

Code for illustration

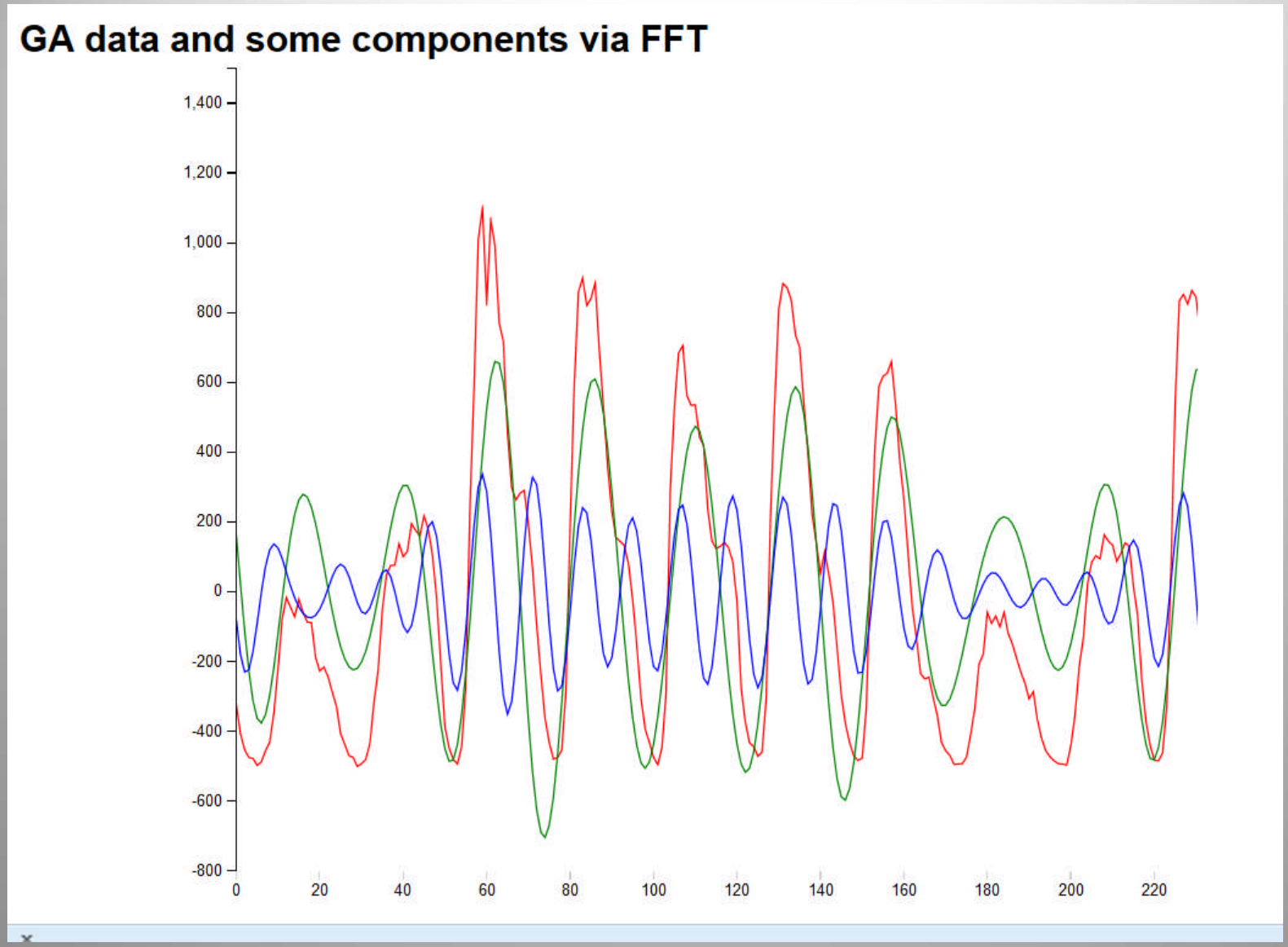
I didn’t go through each line

```
<script>
    d3.json("zero-mean-y.json", function(error, data) {
        plotChart(data); //Load the json data
    });

    function plotChart(data) {
        var w = 800; //width
        var h = 400; // height
        var x = d3.scale.linear().domain([0, ((data.length) - 0) / 24]).range([0, w]); // 24
        var y = d3.scale.linear().domain([-800, 1500]).range([h,0]);

        var xAxis = d3.svg.axis().scale(x).orient("bottom"); //x axis at bottom

        var yAxis = d3.svg.axis().scale(y).orient("left"); // y axis on left

        var line = d3.svg.line().x(function(d, i) {
                    return x(i) / 24; // divide x values by 24 to get hours in days
                }).y(function(d, i) {
                    return y(d); });

        var graph = d3.select("#graph").append("svg:svg").attr("width", w )
                .attr("height", h+ 200 ).append("svg:g")
                .attr("transform", "translate(" + 100 + "," + 100 + ")"); // move chart down

        graph.append("svg:path").attr("d", line(data)); // red line

                graph.append("g").attr("class", "x axis")
                 .attr("transform", "translate(0," + h + ")").call(xAxis);
        graph.append("g").attr("class", "y axis").call(yAxis);

                graph.append("text").attr("class", "x label")
                .attr("text-anchor", "end").attr("x", w - (w / 2))
                .attr("y", h + 45).text("Time in Days");

                graph.append("text").attr("class", "y label")
                .attr("text-anchor", "end").attr("y", -60).attr("x", -60)
                .attr("dy", ".75em").attr("transform", "rotate(-90)")
                .text("Number of Visits");
    }
```

The Desired Result

Filtered data and original data using d3.js

Not finished



http://erushi.com/brightonphp/allseries.html

In summary

- Set up Google service account/key
- Register account with Google analytics
- Get the data and save as JSON
- Process the data with FFT
- Present the data with d3.js

Thank you

Code available at:

http://www.jasonbailey.net/stuff/brighton-php-october-2013-talk/

https://github.com/espeecat/brightonphpfftd3js2013

- Will work on localhost
- remember:
  - Set-up google service account (email address)
  - Download key
  - Allow email address to access GA profile

Twitter @espeecat http://www.jasonbailey.net