Data Source (#1)
Google stores my location history from my Android Phone (https://maps.google.com/locationhistory). I want to analyze my personal data from the last 30 days.

Initial Question (#2)
How frequently is Google polling and storing my location?

Data Inspection (#2)
Google lets you download your data as a KML file that you can open in Google Earth. While I can use the map on their web page or Google Earth to see where I’ve been, it’s not the best way to figure out details of when this information is captured. A quick peek at the raw data looks like this:

It looks like KML is really just XML – after some header stuff, it looks like the raw data is a time stamp, followed by longitude, latitude, and altitude coordinates. The time stamps should allow me to answer my question.
Data Wrangling (#2)

In order to get this into Tableau, the easiest thing I can think of is to convert it to a table. For this, I wrote a quick python script. After a little experimentation with Tableau's date format, here's the script and Tableau interpreting the columns properly:

```python
infol = open('history-67.19-2014.kal', 'r')
outfil = open('history.csv', 'w')

# Write the header.
outfil.write('time,longitude,latitude,alt\n')

for line in infil:
    if line.startswith('\where\n'):
        # Remove whitespace and the tags
        line = line.strip()[:61-7]
        # Extract the data
        date = line[0:18]
        time = line[11:23]
        # Write both separated by a space, followed by a comma
        outfil.write(date + ' ' + time + ',')
    elif line.startswith('\gpx\n'):
        # Remove whitespace and tags
        line = line.strip()[:51-11]
        # Separate the columns
        line = line.split()
        # Write the three values, separated by commas
        outfil.write(',-'.join(line))
        # Write a new line
        outfil.write('\n')

infil.close()
outfil.close()
```

<table>
<thead>
<tr>
<th>Time</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Alt</th>
</tr>
</thead>
<tbody>
<tr>
<td>7/19/14 11:03:53 PM</td>
<td>-111.871127</td>
<td>40.691988</td>
<td>0</td>
</tr>
<tr>
<td>7/19/14 11:08:38 PM</td>
<td>-111.871127</td>
<td>40.691988</td>
<td>0</td>
</tr>
<tr>
<td>7/19/14 11:13:23 PM</td>
<td>-111.871127</td>
<td>40.691988</td>
<td>0</td>
</tr>
<tr>
<td>7/19/14 11:18:08 PM</td>
<td>-111.871127</td>
<td>40.691988</td>
<td>0</td>
</tr>
<tr>
<td>7/19/14 11:21:00 PM</td>
<td>-111.871127</td>
<td>40.691988</td>
<td>0</td>
</tr>
</tbody>
</table>
Initial Exploration
As a sanity check, I made a quick map showing where I’ve been, with the size of each mark encoding the altitude.

Right off the bat, I can see that Google only once properly recorded my altitude (Left: all other readings are zero). I only had my GPS on once this last month (this morning) – could that be why? A quick peek at the underlying data of that big point confirms that it was indeed recorded today (Right).

For fun, I also decided to look at the day of the week, as well as week number:

These shows a couple of things I already know: the marks up by Bear Lake clearly show a camping trip spanning two days (though there is some noise – a mark is on the wrong side of the lake. This is likely because Google uses the cell phone tower location in absence of GPS data). It also shows that I go to Tooele County on weekends (visiting family), and a couple excursions on Thursdays (my wife and I sometimes take Thursday off work for these things).
Now I want to know something about the frequency of each reading – my original question. First, I created a plot of just the number of readings per day (Left). This appears to confirm my hypothesis that Google varies how it samples my location; it takes a different number of readings each day. If I look at the time of day, there’s actually a pattern (Right).

It seems that the number of readings per hour has something to do with when I’m awake. There doesn’t appear to be fine-grain periodicity that we can see at this level. A couple quick checks at the minute and second levels confirms this; if there was a consistent, regular query as to my location (e.g. where am I every 5 minutes?), we should see a significant spike in either of these charts:
A New Question (#3)
This changes my question a little: a specific frequency doesn’t appear to exist. Instead, I’d like to know now what factors affect when a reading is taken?

Here’s a chart showing the count of readings geographically (the one on the right is zoomed in on Salt Lake City):

![Chart showing count of readings geographically](image)

The dots away from my home and the University appear to be disproportionately large, compared to what I know about how much time I’ve spent at each location. Could it be that the more I deviate from my usual locations, the more readings are taken? For this, we need to derive distance from home with this formula:

\[
\text{SQRT}([\text{Latitude}] - 40.6809262) + \text{SQRT}([\text{Longitude}] + 111.8619508)
\]

We also need frequency instead of just the count of readings – I changed my python script to calculate the number of seconds between each reading:

```python
# Start
with importing...

def calculate_distance(latitude, longitude):
    distance = math.sqrt((latitude - 40.6809262)**2 + (longitude + 111.8619508)**2)
    return distance

# Change to calculate frequency...

def main():
    # Read data from file...
    for line in file:
        # Parse data...
        # Calculate distance...
        distance = calculate_distance(latitude, longitude)
        # Calculate frequency...
        frequency = 1 / distance
        # Store the result...

main()
```

And now I can plot the relationship between the two...
Viewed geographically, and considering the known history of the data, it appears that the data is sampled disproportionately when the user is far from home. However, this is only faintly supported in the plot of those two values alone; a rough downward trend exists in the time between samples, but there are clearly other variables involved.